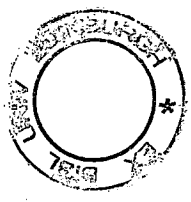
The molecular basis for the regulation of differentiation and maintenance of uncommitted cell populations by PouV proteins

Fella Hammachi

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh
2009



I declare that this thesis has been composed by myself and that the work described in this thesis is my own, except where otherwise stated

Fella Hammachi

Acknowledgments

I would like to thank my supervisor, Dr. Josh Brickman, for giving me a chance to continue my PhD studies in his lab, for this I am very thankfull.

I would like also to thank Dr. Valerie Wilson for her great help during my first two years in the institute.

Big thanks also to Dr.Gillian Morrison for her help and guidance, particularly during the first year of my PhD, and to Dr. Alessandra Livigni for all her friendly help and a fruitful collaboration in the Xenopus experiments.

To all members of our lab, past and present, especially Maurice and Afifah, I thank you all for your help and friendship.

Also, I thank all the ancillary ISCR staff for making things possible, and everyone who offered me reagents, general guidance and help. To all people in the Thursday developmental biology club, thank you for very useful discussions and suggestions.

I also want to thank our collaborators, Minoru Ko and Alexi Sharov at the NIH for undertaking the microarray experiment and the analysis.

To my family, to mum and dad, thank you for your endless support, and encouragement throughout the years.

Finally, I would like to thank my husband for all his endless love, support, and sacrifices throughout the years; I will be forever indebted to you.

Abstract

Oct4 is a member of the Class V POU domain family of transcription factors, and a master regulator of embryonic stem cell pluripotency and lineage commitment in embryos. A number of studies have attempted to define Oct4 targets and have suggested that Oct4 might have a direct role in the suppression of lineage specific gene expression. However, despite a wealth of genomic data, it is unclear whether direct repression of developmental gene expression by Oct4 is relevant to its capacity to suppress differentiation or maintain the pluripotent state. In this thesis I have established correlations between PouV proteins molecular function and their capacity to support ES cell growth, both based on an analysis of structure function and the global activity of these proteins in transcription.

In the first part of this thesis, I have directly tested the capacity of repression by PouV proteins to regulate ES cell self-renewal and embryonic differentiation by asking whether activator or repressor fusion PouV proteins are capable of suppressing differentiation and maintaining ES cells self-renewal. I employed a series of reiterated modular activation and repression domains fused to either Oct4 or a functional Xenopus homologue of Oct4, Xlpou91, to generate global activators or repressors of Oct4 target gene expression. Expression of Oct4 or Xlpou91 activator forms (Oct4λVP2, Xlpou91λVP2) in both Xenopus embryos and ES cells appears to suppress differentiation, while expression of their repressor forms (Oct4λEnR, Xlpou91λEnR) induces it. Moreover, activator fusions to either Oct4, or Xlpou91 maintain murine ES cells in the absence of endogenous Oct4. Oct4λVP2 cells express pluripotency genes and are resistant to differentiation in the presence of high levels of Oct4 expression or upon LIF withdrawal, a phenotype that seems correlates with Nanog levels. More importantly, profiling of Oct4λVP2 expressing cells indicates that the same set of genes that are downregulated in response to Oct4 depletion are upregulated in response to Oct4λVP2, indicating that positive regulation of the pluripotency network appears to be the essential activity of Oct4.

Although, activation of transcription by PouV proteins seams to be sufficient for maintaining self-renewal and blocking differentiation in both ES cells

and Xenopus embryo, previous study carried out in our laboratory revealed that while PouV proteins appear to have equivalent activity in transient reporters assays, they vary greatly in their ability to maintain ES cell self-renewal in place of the murine Oct4. These findings suggest that the difference in their activities might depend on the presence or the absence of specific protein-protein interaction motifs required for Oct4 to recognize complex promoters in vivo.

In the second part of this thesis, I identified PouV protein domains responsible for maintaining the Oct4-dependent undifferentiated phenotype in ES cells by exploiting the sequence similarities and the functional differences between these PouV proteins. I constructed a series of chimeric proteins by replacing domains from the two proteins (Xenopus Xlpou25, and Zebrafish DrPou2) that have only low levels of Oct4-like activity, with those derived from the Xenopus Xlpou91 protein, a highly related protein with an equivalent activity to mouse Oct4, and tested their ability to rescue Oct4 knockdown in ES cells. These chimeric proteins generally activated transcription equally in transient reporter assays, and differed in their ability to maintain ES cell self-renewal in place of Oct4. Differences in the phenotypes supported by these chimeras suggest that two separable activities have been lost by these non-functional PouV proteins in evolution. A quantitative capacity to support clonal alkaline phosphatase positive growth that is dependent on an Xlpou91/Oct4-like Carboxyl terminal region of the protein, and a capacity to maintain a normal ES cell colony morphology, which seems to depend on an Xlpou91/Oct4-like POU domain. In addition, through the analysis of new Zebrafish PouV sequences, I have identified a single point mutation, that does not effect the protein ability to activate transcription from reporter genes, but can completely eliminate its ES cell colony forming potential.

Taken together my data suggests that the correct recognition of endogenous targets by certain PouV family members requires interactions within both the POU domain and Carboxyl terminus. While there maybe examples of targets that are repressed by PouV containing protein complexes, repression by Oct4 is not required for the maintenance of the undifferentiated state. Rather the predominant role of the PouV proteins is to activate a network of gene expression that blocks differentiation in multiple lineages, in both ES cells and embryos.

Abbreviations

a.a	amino acid
bp	base pair
BSA	Bovine Serum Albumin
°C	Degrees centigrade
$CaCl_2$	Calcium Chloride
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
cm^2	Square centimetre
CO2	Carbon dioxide
C-TD	Carboxy- terminal domain
DAPI	4',6-diamidino-2-phenylindole
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylenediamine tetra-acetate
ES cells	Embryonic Stem cells
EtOH	Ethanol
FCS	Fetal Calf Serum
g	Grams
GMEM	Glasgow's modified Eagle's medium
H2O	Water
h	Hour
IF	Immunofluorescence
IRES	internal ribosome entry site
kb	Kilobase
kDa	KiloDaltons
l	Litre
LB	Luria-Broth
LIF	Leukemia inhibitory factor
m	Milli
µ	Micro
M	Molar
MgCl2	Magnesium chloride
min	Minutes
n	Nano
NaCl	Sodium choride
NaOAc	Sodium acetate
NaOH	Sodium hydroxide
N-TD	amino terminal domain
N2B27	Neural basal medium
OD	Optical Density
Ori	Origin of replication
P	Pico
pA	polyadenilation
Pac	puromycin acetyl transferase
PBS	Phosphate-buffered saline

PBST	Phosphate-buffered saline-Tween20 soltion
PCR	Polymerase Chain Reaction
mRNA	messenger Ribonucleic acid
rpm	revolutions per minute
RT	room temperature
RT-QPCR	real-time quantitative polymerase chain reaction
Sec	Second
SDS	sodium dodecyl sulfate
SV40	Simian vacuolating virus
TAE	Tris-acetate-EDTA
Tc	Tetracyclin
UV	Ultraviolet
V	Voltage
v/v	Volume/ Volume
w/v	Weight/Volume
WB	Western Blotting
3XFlag	3 copies of Flag tag
7AAD	7-Amino-actinomycin D

Table of Contents

Chapter 2: Materials and Methods

Chapter3: Oct4 activation function is sufficient to block lineage commitment and maintain ES cell self-renewal

Chapter4: Mapping functional domains of Oct4 homologues responsible for ES self-renewal

List of figures

Figure 4.20-The inability of Zebrafish DrPou2(A) protein to rescue ES cells self-renewal reveals the crucial role of POU domain of the PouV proteins in the process

Figure 5.1-Effect of Oct4λVP2 on the core pluripotency network in mouse embryonic stem cells

List of tables

CHAPTER 1

General Introduction

1. Introduction

1.1. Early mouse development

Following fertilization, the zygote undergoes multiple cell divisions, a process known as cleavage. These divisions lead to the formation of a solid mass of cells called the morula (16 cell stage) and culminate in forming the blastula at embryonic day 3.5 (E3.5) of mouse development. At E4.5, the three primary lineages (tissues) of the pre-implantation mouse embryo: trophectoderm (TE), epiblast and primitive endoderm (PE) are established. TE will give rise to the trophobalst and extraembryonic ectoderm (ExE), the epiblast is a pluripotent tissue, which will differentiate to form all the germ layers of the embryo proper, and the PE is the tissue that parietal and visceral endoderm originate from (Figure 1.1a).

The molecular cascade regulating these early lineage specification and segregation is not well understood. However, recent study started to unveil some key factors involved in these processes. Caudal-type homeobox protein 2 (Cdx2) and octamer binding transcription factor 4 (Oct4 also called Oct3/4 and Pou5f1) direct the first cell fate choice to form TE and inner cell mass (ICM), respectively (Niwa et al., 2005). Cdx2 expression starts around the 8-cell stage in all blastomeres, which overlaps with Oct4 expression (Figure 1.1b). Interestingly, Cdx2 expression becomes restricted and upregulated in the outer cells of the morula (16 cell stage). At embryonic day 3.5, the TE and the ICM, the first two lineages of the embryo are generated (Figure 1.1a). The TE is the outer layer of the blastocyst, and is composed of larger cells exclusively expressing Cdx2, whereas the ICM is composed of the smaller inner cells expressing Oct4 (Figure 1.1b). The reciprocal inhibition between Cdx2 and Oct4 coupled with their autoregulatory properties maintain their exclusive expression to their respective lineages. The initial localization of Cdx2 to the outer cells is regulated by TEA-domain family member 4 (TEAD4) through the sub-cellular localisation of its co-activator Yap (Yes-associated protein) (Nishioka et al., 2009). Yap is localised in the nuclei of outer cells of the morula, where it binds to TEAD4 and induces Cdx2 expression, while in inner cells Yap is cytoplamsic

resulting in TEAD4 inactivity (Nishioka et al., 2009). It has been reported that the sub-cellular localization of Yap is regulated by large tumour suppressor, homolog 1-(Lats1-) mediated phosphorylation downstream of Hippo-induced cell signalling, which in turn might be influenced by the degree of cell contact, cell polarization or the presence of an exposed apical surface (Nishioka et al., 2009).

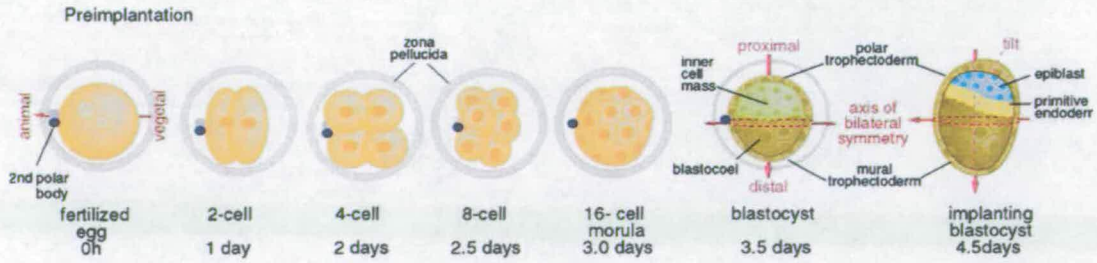
The second cell fate decision segragates the PE and the epiblast lineages from the ICM at the blastocyst stage (E3.5) (Figure 1.1b). At the early blastosyst stage, the ICM cells show a mosaic expression pattern of the transcription factors, Nanog and GATA-binding factor 6 (GATA6). Cells expressing GATA6 are organised to the distal surface of the ICM to give rise to the primitive endoderm, while, Nanog-expressing cells generate the pluripotent epiblast (Figure 1.1b). It has been reported that active signalling through growth-factor-receptor-bound protein 2 (Grb2), a mediator of receptor tyrosine kinase–Ras–mitogen-activated protein kinase (MAPK) signalling pathway (Chazaud et al., 2006), is necessary for the initiation of the primitive endoderm gene expression (Chazaud et al., 2006; Cheng et al., 1998; Fujikura et al., 2002). Moreover, FGF signalling might also be implicated in the primitive endoderm development (Chazaud et al., 2006).

After implantation into the uterine wall, a cavity forms within the epiblast causing it to acquire a cup-like shape along the proximo-distal (P-D) axis (Figure 1.1c). This P-D axis of the pre-gastrulation embryo is induced by a proximal-to-distal gradient of Nodal signalling emanating from reciprocal signalling between VE, ExE and the epiblast. Up to this point the conceptus is bilaterally symmetrical, but at E6.0 the first morphological indication of asymmetry becomes apparent on the future posterior side with the emergence of the primitive streak, marking the initiation of gastrulation (Figure 1.1c) (Beddington and Robertson, 1999).

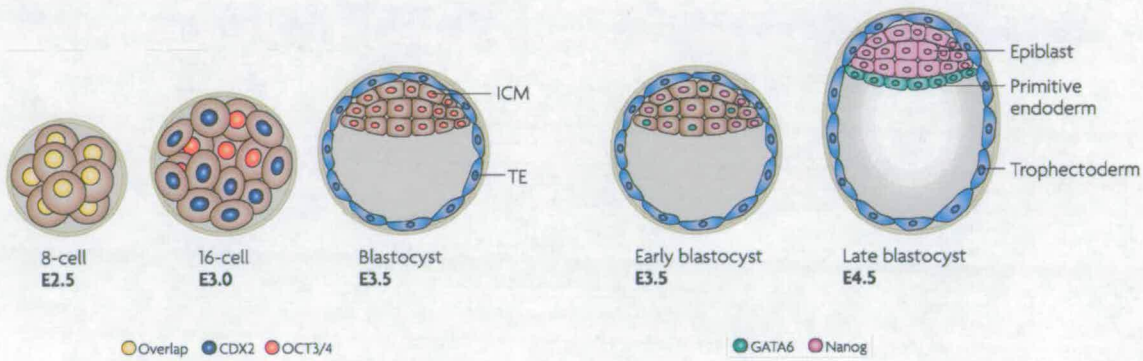
Gastrulation is a complex and co-ordinated series of cellular movements that leads to the formation of the three embryonic germ layers: ectoderm, mesoderm and definitive endoderm. This complex process is controlled by graded signals differentially converging on transcriptional regulators to confer spatio-temporal information essential for cell fate commitment. At the start of gastrulation (E6.5),

cells of the epiblast migrate towards the posterior pole of the embryo and ingress at the primitive streak and undergo an epithelial mesenchymal transition to form the nascent embryonic mesoderm and extraembryonic mesoderm. As the primitive streak extends to the distal end of the embryo, the cells in the anterior primitive streak go to form the definitive endoderm. The definitive endoderm appears on the surface of the embryo and gradually replaces the visceral endoderm. The cells that are left in the epiblast by the end of gastrulation form the neurectoderm (Beddington and Robertson, 1999). Figure 1.1c illustrates the development of the early mouse embryo until the beginning of gastrulation.

(a)



(b)



(c)

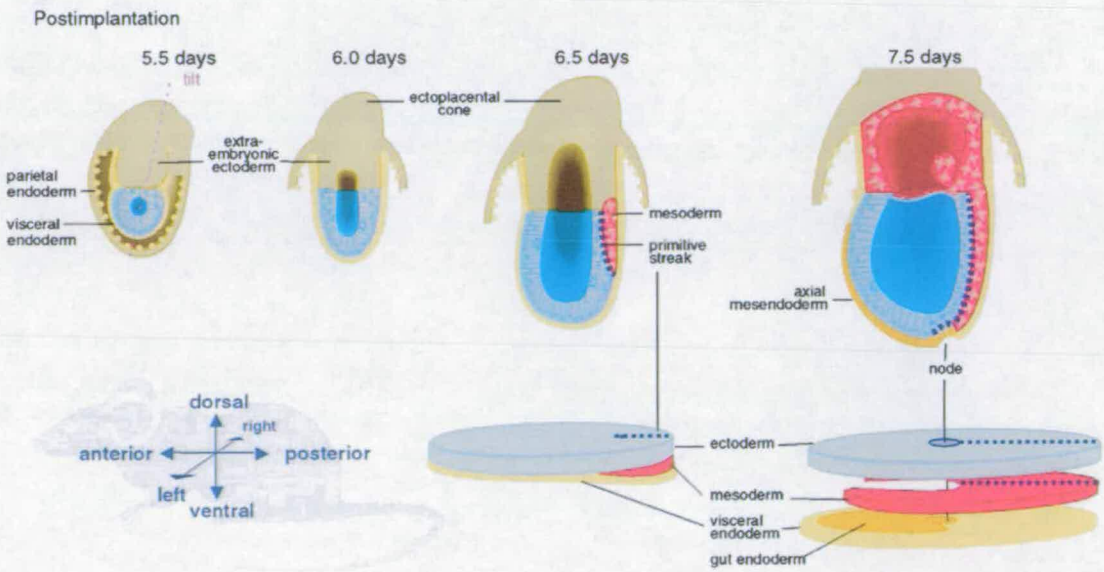


Figure 1.1- Early mouse embryogenesis

(a) The stages of pre-implantation development from fertilization to blatsocyst formation showing the early lineages of the mouse embryo.
(b) The molecular cascades governing the first and second lineage specification in the pre-implantation embryo.
(c) The initiation of gastrulation in the post-implantation embryo.
Adopted from ((Beddington and Robertson, 1999); (Arnold and Robertson, 2009)

1.2. Derivation of mouse embryonic stem (ES) cells

Embryonic stem cells are karyotypically normal self-renewing pluripotent cell lines derived originally from the mammalian blastocyst ICM (Brook and Gardner, 1997; Evans and Kaufman, 1981; Martin, 1981). In addition to ES cells a number of other multipotent cell lines have been derived from the murine pre-implantation embryo including trophoblast stem cells (TS), and extraembryonic endoderm stem cells (Xen) (Kunath et al., 2005; Tanaka et al., 1998) (Figure 1.2). All three of these cell lines are generated by culturing blastocysts under different conditions, but only ES cells are pluripotent. Moreover, the conditions for human ES cell derivation are different, and despite having been derived from blastocysts, they have different morphology and growth requirements (Thomson et al., 1998).

The basis for these differences became clearer with the isolation of two other types of embryo-derived stem cells; the FAB-S (fibroblast growth factor/activin/BIO-stem) and the Epi-S (epiblast-derived stem) cells (Figure 1.2). FAB-S cells are blastocyst-derived stem cells that have been shown to be "partially" pluripotent (Chou et al., 2008), whereas, Epi-S are epiblast-derived multipotent cells (Brons et al., 2007; Do and Scholer, 2009; Tesar et al., 2007). Analysis of the Epi-S and FAB-S cells showed that they have a restricted developmental potential and lack some of the pluripotency markers maintained in ES cells (Brons et al., 2007; Chou et al., 2008; Tesar et al., 2007). Interestingly, the expression profile of human ES cells is closer to these cell lines than they are to mouse ES cells, suggesting that human ES cells may represent a later stage of development (Do and Scholer, 2009).

As the definition of pluripotency is based on the ability of cells to colonize the germ line in chimaeric embryos, it is difficult to assess pluripotency in Epi-S and FAB-S cells, as their potency can only be assayed by teratoma formation and their capacity to differentiate towards specific lineages *in vitro*. However, in addition to ES cells there are other pluripotent cell lines that can be derived either from germ cells or through reprogramming of adult somatic cells. Embryonic germ cells (EG) are derived from primordial germ cells, are pluripotent cells that can be cultured indefinitely *in vitro* (Matsui et al., 1992; Resnick et al., 1992) (Figure 1.2) and have the potential to differentiate into all the cell types of the body (Labosky et al., 1994;

Stewart et al., 1994). Induced pluripotent stem cells (iPS) are pluripotent cells that were artificially derived from non-pluripotent adult somatic cells by ectopically expressing certain combination of ES cell specific transcription factors (Takahashi and Yamanaka, 2006) (Figure1.2).

The pluripotency of ES, EG and iPS cells coupled with their indefinite self-renewal capacity make them an invaluable source for tissue engineering and regenerative medicine. Consequently it is crucial to understand the molecular mechanisms responsible for pluripotency, self-renewal and differentiation for reliable manipulation of these cells *in vitro*.

One of the central factors common to all of these cell types is the transcription factor Oct4. This thesis focuses on the activity of Oct4 in maintaining self-renewal and pluripotency.

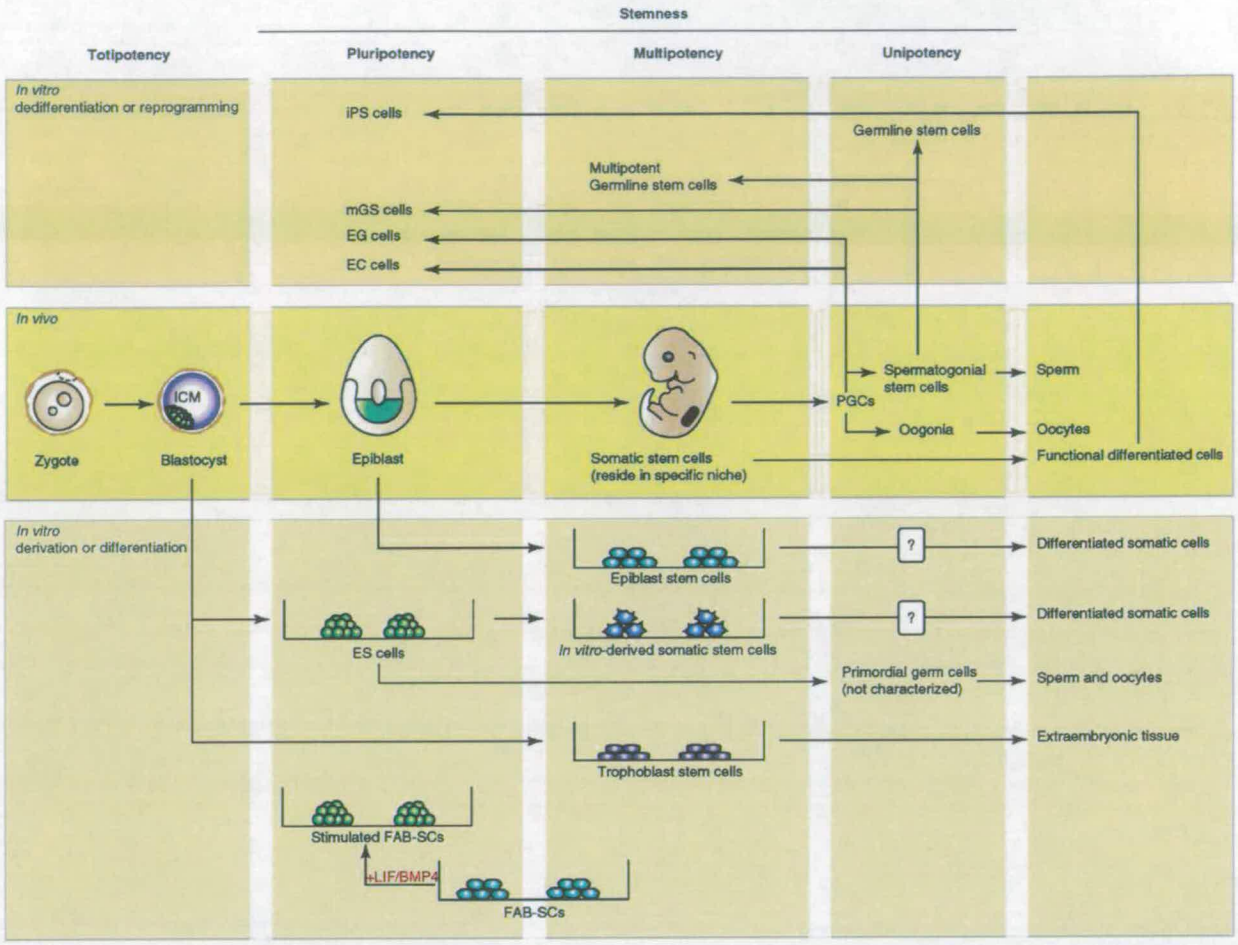


Figure. 1.2- Embryo-derived cells and their potency.

The zygote is totipotent, while cells derived from the ICM (ES cells) or the germ cells (EG, EC and mGS cells) are pluripotent. iPS cells are also pluripotent. Epiblast-derived progenitor cells are multipotent can be reprogrammed to pluripotency by addition of external factors or forced expression of iPS factors.
Adopted from (Do and Scholer, 2009)

1.3. Molecular basis of pluripotency

A number of intrinsic (transcription factors) and extrinsic (cytokines, small molecules) contribute to maintaining ES cells in a pluripotent state and blocking differentiation.

1.3.1. Extrinsic factors

Originally, propagation of ES cells in culture required the presence of serum and fibroblast feeder layer (Smith and Hooper, 1983). The requirement for fibroblasts-conditioned medium was replaced by Buffalo rat liver-conditioned medium (Smith and Hooper, 1987), and subsequent fractionation of the medium revealed leukaemia inhibitory factor (LIF) as the active component (Smith et al., 1988; Williams et al., 1988). LIF was the first extrinsic factor shown to be required for the maintenance of ES cell pluripotent state and allowed their propagation without feeders in the presence of serum. The component in serum required for ES cell maintenance was later discovered to be Bmp4 (Ying et al., 2003).

1.3.1.1. LIF/Stat3 pathway

LIF is a member of the IL6 (Interleukin 6) family of cytokines. It is a secreted chemokine that acts through binding to transmembrane tyrosine receptor kinases (TRKs). The receptor for LIF is a heterodimer composed of glycoprotein 130 (gp130) and LIF receptor (LIFR) (Ernst and Jenkins, 2004). The intracellular fraction of the receptor interacts constitutively with the inactive form of the tyrosine kinase, janus kinase (JAK) (Figure 1.3a). Upon LIF binding, JAK phosphorylates both gp130 and LIFR, which in turn recruit the signal transducers and activators of transcription (Stats). Stats are recruited by the **S**rc **H**omology 2 (SH2) domains of gp130 and LIFR to the cytoplasmic membrane (Stahl et al., 1995), and then get phosphorylated by JAK. This phosphorylation of Stats causes them to dimerise and then translocate into the nucleus to activate transcription (Figure 1.3a)

(Auernhammer et al., 2000)

It was reported that the ability of LIF to maintain mouse ES cell self-renewal is dependant on Stat3 activation (Niwa et al., 1998), and that its activation is sufficient for maintaining mouse ES cell self-renewal in the presence of foetal calf serum (Matsuda et al., 1999). Interestingly, the LIF receptor is not specifically required as alternative pathways that activate Stat3 are also able to maintain ES cells undifferentiated. Provision of IL6 together with soluble IL6 receptor can support ES cell derivation and maintenance in the absence of LIF (Nichols et al., 1994; Yoshida et al., 1994).

Despite its central role in maintaining ES cell self-renewal, the identity of the crucial LIF targets remains largely obscure. The identification of these downstream targets is a crucial component of understanding how pluripotency and ES cell self-renewal is maintained. To identify targets, Sekkai et al (2005) compared ES cell global gene expression in response to LIF withdrawal over time with and without expression of the dominant negative Stat3 mutant (Sekkai et al., 2005). Expression of *Lefty1, Id1, Id2, Esg-1* and *Aes1* were all shown to correlate with Stat3 gene expression. *Aes1*, a groucho-like protein, and *Esg-1* were also identified as potential direct targets of Stat3 (Sekkai et al., 2005; Tanaka et al., 2002). Cartwright et al (2005) found that c-Myc expression in ES cells was also dependent on LIF/Stat3 signalling. c-Myc expressing ES cells are able to self-renew in a LIF independent manner and a dominant negative form of c-Myc induces differentiation (Cartwright et al., 2005). c-Myc is an oncogene that is overexpressed in different cancers, is a potent inducer of proliferation, and was identified as one of the key factors responsible for the generation of iPS cells (see below).

In addition to the Stat3 pathway, LIF interacts with other signalling pathways in ES cells. Paradoxically Stat3 stimulates differentiation through Ras-MAPK signalling pathway (Burdon et al., 1999), while normally acting through PI3K (phosphoinositide-3 kinase) signalling pathway to suppress the differentiation–promoting activity of ERK (Paling et al., 2004).

Although LIF signalling is required for the derivation and maintenance of ES cells, it

is not essential for mouse blastocyst development. Despite the fact that *Lif*, *Lifr*, *gp130* are co-expressed in the early mouse embryo (Nichols et al., 1996), ablation of any of them, does not affect normal blastocyst development (Stewart et al., 1992; Ware et al., 1995). The LIF signalling pathway was only required for blastocyst development during diapause where development becomes slower (Nichols et al., 2001). This suggested that LIF-gp130 signalling is essential for the prolonged maintenance of the epiblast *in vivo*, and that this adaptive physiological mechanism provides the basis for ES cell requirement of LIF-gp130 signalling (Nichols et al., 2001).

Interestingly, human and monkey ES cells maintain their pluripotency in a LIF/Stat3-independent manner despite the presence of the required components of the LIF/Stat3 pathway in these cells (Daheron et al., 2004; Humphrey et al., 2004).

1.3.1.2. Bmp/SMAD pathway

Bone morphogenetic 4 (Bmp4), a member of the transforming growth factor (TGFβ) superfamily, was shown to substitute for the requirement of serum during ES cell propagation and derivation (Ying et al., 2003). While the TGF-ß family is large, only Bmp4, Bmp2 and GDF6 have this capacity (Ying et al., 2003). These ligands bind to hetrodimeric complexes of type I and the ligand-specific type II receptors (Figure 1.3b). The type II receptor is a serine threonine kinase that recruits and phosphorylates the type I receptor upon ligand binding. The type I receptor then phosphorylates a receptor-regulated SMAD (RSmad,). For Bmp signalling the RSMADs are SMAD 1, 5 and 8 (as opposed to 2 and 3 for Nodal/TGF-beta signalling), which in turn heterodimerise with SMAD4, a cooperating SMAD (C-SMADs) (See Figure 1.3b). These SMAD heterodimers translocate to the nucleus, interact with specific transcription factors and stimulate transcription (Hill, 2009).

Ying et al (2003) reported that Bmp signalling maintains ES cell self-renewal by blocking neural differentiation through the activation of inhibitors of the pro-neural transcription factors, dominant negative bHLH proteins that inhibit the DNA by forming inhibitory heterodimers. Stimulation of Smad1, 5, and 8 in ES cells

maintains the expression of a number of these inhibitory (Id) genes and exogenous expression of Id proteins (*Id1*, *Id2*, or *Id3*) bypasses the requirement for Bmp signalling in the maintenance of ES cell self- renewal (Ying et al., 2003).
Although the downstream targets of the Id proteins in ES cells are not yet identified, the Id proteins may be repressing neural-specific (bHLH) transcription factors such as Mash1 and Neurogenin 2 (Ying et al., 2003), or non bHLH proteins such as Pax factors (Norton, 2000). It was also proposed that Bmp signalling maintains ES cell self-renewal by inhibiting the ERK and p38 MAPK pathways (Qi et al., 2004).

The Bmp4 and LIF pathways co-ordinate their functions to block both neuronal and non-neuronal differentiation. Thus, in the presence of LIF alone, ES cells differentiate towards the neuronal lineage, however, addition of Bmp4 or activation of Id proteins blocks this neuronal differentiation. Similarly, presence of Bmp4 alone is not sufficient to block mesodermal and endodermal differentiation that is inhibited by LIF. Moreover, Stat3 and SMAD1 are known to form a complex in neuropithelial cells, and cooperatively regulate transcription (Nakashima et al., 1999). This complex was also detected in mouse ES cells (Ying et al., 2003). The formation of the complex is likely to be dose-dependent. The precise balance between activated Stat3 and Smad1 is important for the maintenance of self-renewal as the overexpression of SMAD 1/4 can induce non-neural differentiation even in the presence of LIF (Ying et al., 2003). The interaction between the LIF and Bmp pathways is supported by genetic studies in the mouse. The ablation of *Bmp2* renders the uterine environment incapable of supporting embryonic development after implantation (Lee et al., 2007b), which is reminiscent of the *Lif* knockout mouse phenotype (Stewart et al., 1992).

As with LIF, Bmp signalling has a very different role in human ES cells. Bmp4 induces differentiation towards mesoderm and ectoderm, whereas Bmp2 addition directs ES cells towards the extraembryonic endodermal lineage (Schuldiner et al., 2000). Moreover, the antagonism of Bmp signalling by Noggin in the presence of basic FGF (bFGF) maintains human ES cell self-renewal in the absence of both serum and feeder cells (Xu et al., 2005).

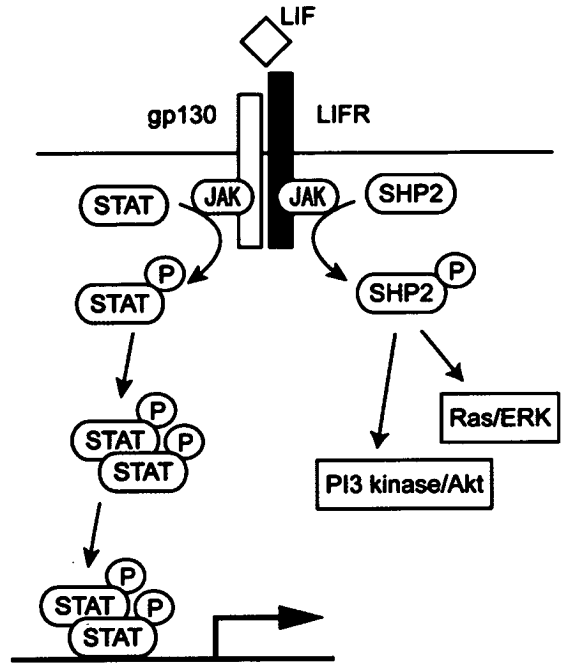
1.3.1.3. Wnt/β-catenin pathway

Wingless-related MMTV integration site proteins (Wnts) are secreted glycoproteins that are involved in tissue differentiation and organogenesis (Cadigan and Nusse, 1997). Wnts bind their membrane receptors, Frizzled and low density lipoprotein-related receptor (LRP) and signal through several pathways. The best characterized or so called "canonical pathway" signals through β-catenin-mediated transcriptional program regulating cell fates (Logan and Nusse, 2004), while there are also β-catenin independent pathways that signal through calcium flux, JNK, and G proteins (Veeman et al., 2003).

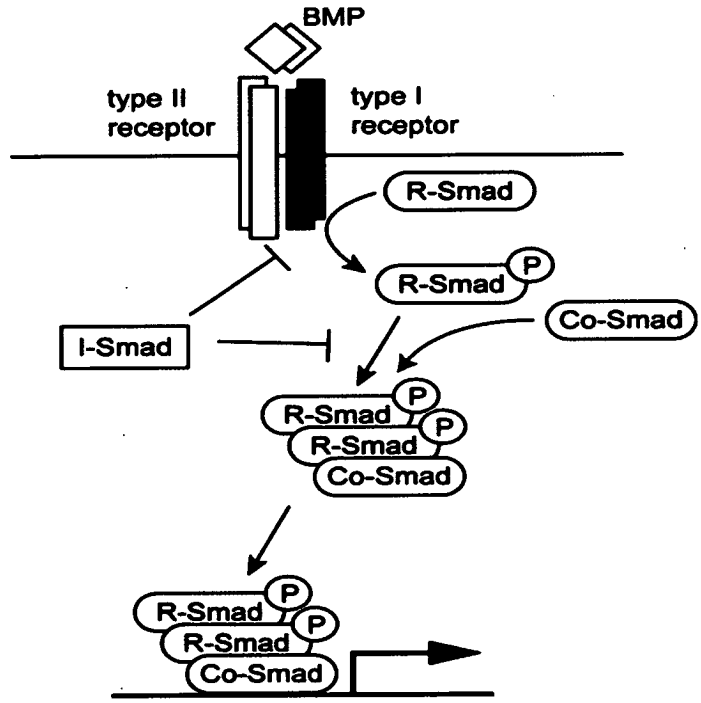
The canonical pathway employs β-catenin as a mediator of cytoplamsic signalling and transcriptional activation of gene expression (Fig 1.3c). In the absence of Wnt signalling, β-catenin is phosphorylated by a protein complex consisting of glycogen synthase kinase (GSK3β), adenomatous polyposis coli gene (APC) and Axin (Figure 1.3c). The phosphorylated β-catenin then undergoes ubiquitin-mediated proteolysis. However, Wnt binding to its receptor leads to activation of Disheveld (Dsh), which then inhibits GSK3β resulting in the stability of β-catenin and its accumulation in the cytoplasm (Figure 1.3c). The accumulated β-catenin enters the nucleus and binds the T cell factors (TCFs) transcription factors converting them from repressors to activators of gene transcription (Figure 1.3c).

Wnt signalling has been shown to play a role both in suppressing differentiation and promoting germ layer induction (Grigoryan et al., 2008). The canonical Wnt pathway plays a vital role in the formation of the anterior-posterior (A-P) axis, and germ layer differentiation, especially mesoderm induction (Huelsken and Birchmeier, 2001; Niehrs, 1999). Canonical Wnt signalling also suppresses neural differentiation as *Wnt3a* (-/-) embryos have no primitive streak and exhibit ectopic neural tube formation in its place (Yoshikawa et al., 1997).

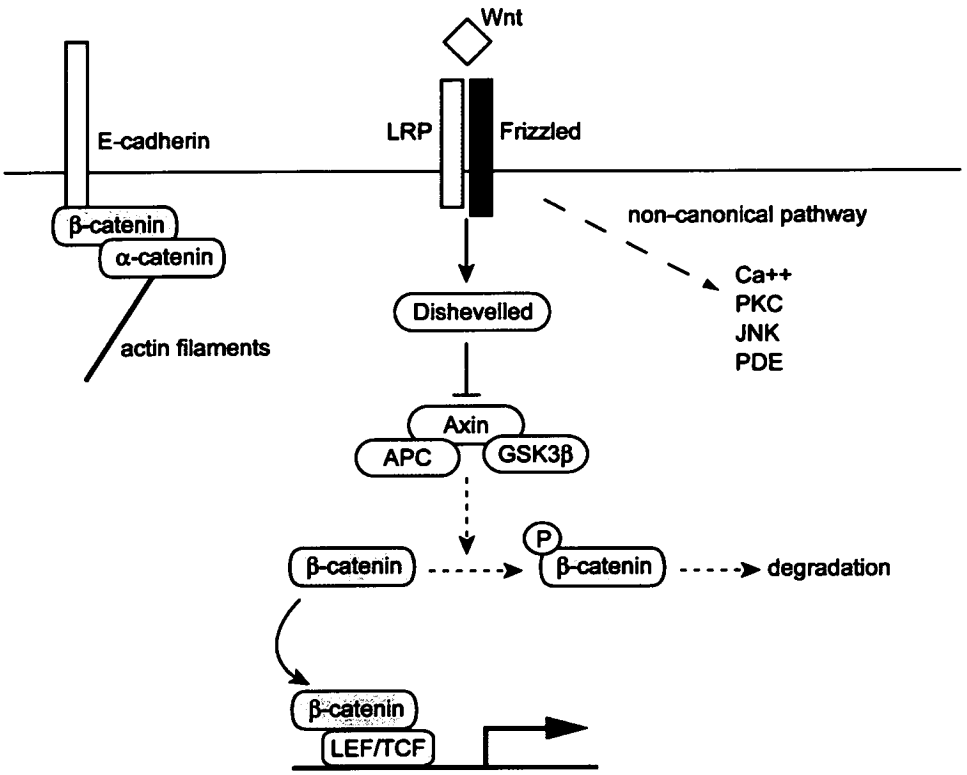
(a)


LIF
gp130
LIFR
STAT
JAK
JAK
SHP2
P
STAT
P
SHP2
P
STAT
P
STAT
STAT
Ras/ERK
PI3 kinase/Akt
P
STAT
P
STAT

(b)


BMP
type II
receptor
type I
receptor
R-Smad
R-Smad
P
I-Smad
Co-Smad
R-Smad
P
R-Smad
P
Co-Smad
R-Smad
P
R-Smad
P
Co-Smad

(c)



(d)

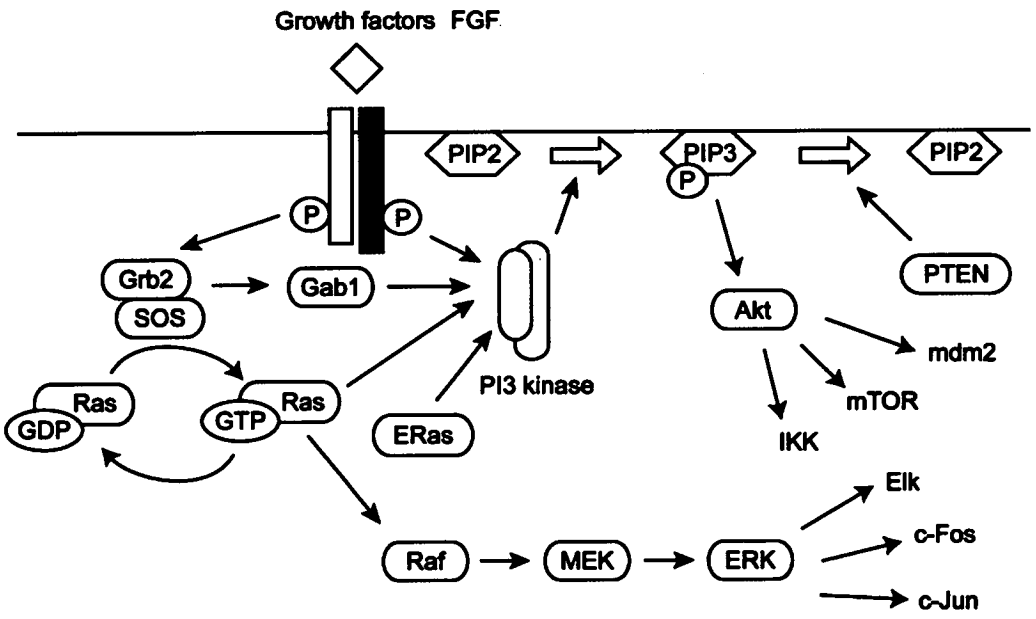


Figure 1.3-The external pathways governing pluripotency and self-renewal

(a) The LIF-Stat3 pathway.
(b) The Bmp-SMAD pathway.
(c) The Wnt/ß-catenin pathway.
(d) The FGF pathway.
Adopted from (Okita and Yamanaka, 2006)

pathway in ES cell pluripotency. Sato et al (2004) reported that Frzd5 is highly enriched in self-renewing ES cells, and that Wnt signalling is endogenously activated in ES cells and only down-regulated upon differentiation (Sato et al., 2004). In contrast to LIF and Bmp signalling pathways, Wnt/β-catenin signalling appears to have a similar role in regulating differentiation in both mouse and human ES cells (Kielman et al., 2002; Sato et al., 2004). The inhibition of GSK-3b activity with either 6-bromoindirubin-3'-oxime (BIO) (Sato et al., 2004) or Chiron (Ying et al., 2008) enhanced the self-renewal of ES cells in undifferentiated state. Furthermore, Wnt3a was implicated in the maintenance of mES and hES cells in undifferentiated state, and required to sustain the expression of pluripotency markers Oct4 and Nanog (Ogawa et al., 2006; Singla et al., 2006). Additionally, the recent discovery of Tcf3; a transcriptional co-binder of β-catenin, as a key factor of the core regulatory circuitry of ES cells, confirms the involvement of this pathway in maintaining ES cell self-renewal (Cole et al., 2008).

1.3.1.4. PI3K pathway

Phosphoinositide kinase (PI3K) is a kinase that catalyses the phosphorylation of lipids at the cell membrane. These lipids are a type of phosphoinositide (PIs) known as phosphotadylinositol phosphates (PIPs) (Figure 1.3d). PI3Ks are divided into three classes depending on substrate specificity and sequence homology (Takahashi et al., 2005). Class 1A of PI3K consist of two subunits: a regulatory, and a catalytic subunit. These subunits come in several isoforms.

Class 1A of PI3Ks is activated by various growth factor receptors such as FGF, EGF, and PDGF (see Figure 1.3d). Activation of PI3K can be achieved via a number of distinct mechanism(s) however, the principle activation occurs via Grb2 (growth factor receptor binding protein 2). Upon binding of growth factors to their receptors, the receptors and their associated proteins are auto-phosphorylated. The adaptor protein Grb2 binds to the phosphorylated receptor and activates the Ras/ERK pathway that in turn activates PI3K (Figure 1.3d). Active Grb2 also binds and phosphorylates Gab1, which can directly activate the PI3K pathway. PI3K can also be activated via direct binding of the PI3K regulatory subunit to the phosphorylated

tyrosine residue of the RTK or GTP-bound Ras (i.e. active Ras) (Takahashi et al., 2005).

The active PI3K localizes at the plasma membrane, and catalyses the phosphorylation of inositol phospholipids (PIP2) generating phosphatidylinositol3,4,5-tris-phosphate (PIP3). These PIP3s are the ligands for the pleckstrin homology (PH) domains of several signal transducers including phosphoinositide-dependent kinase1 (PDK1) and protein kinase B (PKB/Akt). PDK1 and Akt binding to PIP3s leads to their co-localisation to the membrane and their phosphorylation. Active Akt phosphorylates many substrates including mammalian target of rapomycin (mTOR), BAD, mdm2, IKK and the FOXO transcription factors (Takahashi et al., 2005).

This pathway has multiple auto-regulatory loops. Akt and mTOR regulate each other activity through phosphorylation. The pathway is negatively regulated by the p53 target, PTEN (phosphatase and tensin homologue deleted on chromosome ten). PTEN can remove phosphates from PIP3s thereby inhibiting Akt activation. Akt in turn regulates PTEN via p53 degradation. Moreover, Akt-dependent phosphorylation MDM2 induces p53 degradation. As p53 normally activates PTEN, its degradation results in reduced PTEN expression (Stambolic et al., 2001).

The PI3K signalling pathway is involved in proliferation and inhibition of apoptosis during embryogenesis. Genetic ablation of multiple components of the pathway supports a role in the early embryo development, and self-renewal. Expression of catalytic and regulatory subunits of PI3K class Ia starts at the one cell stage of embryo development (Lu et al., 2004; Riley et al., 2005), and the signalling emanating from PI3K is required for pre-implantation development as the inhibition of the PI3K pathway leads to the induction of apoptosis in both murine blastocysts and the trophoblast stem cells (Lu et al., 2004; Riley et al., 2006; Riley et al., 2005). Moreover, the deletion of the gene encoding the PI3K catalytic subunit p110β leads to embryonic lethality at the blastocyst stage (Bi et al., 2002), whereas deleting the p110α catalytic subunit causes mouse embryos to die between E9.5 to E10.5 as a

result of a proliferative defect during gastrulation (Bi et al., 1999). Disruption of other components of the PI3K such as the regulatory subunit genes p85α and p85β also results in lethality at E12.5 (Brachmann et al., 2005).

In mouse ES cells, inhibition of PI3K activity suppressed their transition from the G1 to the S phase, and lowered their proliferation rate (Jirmanova et al., 2002). Similarly, *Pten* null ES cells show an increased rate of cell proliferation and viability (Sun et al., 1999). Moreover, suppression of the PI3K activity appears to reduce the ability of LIF to maintain mouse ES cell self-renewal by increasing the LIF induced phosphorylation of the pro-differentiation signalling kinase extracellular signal-regulated kinase (ERK) (Paling et al., 2004).

1.3.1.5. Ras/Raf/ERK pathway

Ras protein belongs to the low-molecular weight G-proteins super-family. Ras signalling can be activated by binding to many receptor tyrosine kinases such as FGF, EGF, PDFG growth factor receptors (Figure 1.3d). Upon binding of growth factors to their receptors, the receptors and their associated proteins are auto-phosphorylated. The SH2 domain-containing tyrosine phosphatase 2 (SHP2), and Grb2 bind to the phosphorylated receptor (see Figure 1.3d). The binding of these adaptor proteins to the phosphorylated receptor leads to the recruitment of the guanine-nucleotide-exchange factor: son of sevenless (SOS), which in turn catalyses the transformation of the Ras from its inactive GDP-bound state to a GTP-bound active state (Okita and Yamanaka, 2006). Active Ras leads to the phosphorylation and activation of the Raf kinases, which activate ERK (Figure 1.3d). The ERK pathway phosphorylates multiple proteins essential for cell cycle progression, survival and differentiation (Okita and Yamanaka, 2006).

Several studies reported that the Ras/ERK pathway promote differentiation and antagonizes ES cell self-renewal. Burdon and colleagues (1999) showed that inhibiting the ERK signalling with the pharmacological inhibitor PD98059 leads to increased efficiency of ES cell derivation from blastocysts (Burdon et al., 1999). This was later confirmed by Batle-morera and colleagues (Batlle-Morera et al.,

2008). The authors showed that ERK inhibition within the epiblast is important to maintain pluripotency, and demonstrated that ERK inhibition is required for the derivation of ES cells from recalcitrant strains (such as C57BL/6) (Batlle-Morera et al., 2008).

A role for the Ras/ERK pathway in promoting ES cell differentiation was also reinforced by the fact that *Grb2*-deficient ES cells are unable to differentiate towards the endodermal lineage (Cheng et al., 1998). Similarly, ectopic expression of the active form of hRas in human ES cells directs ES cells towards the primitive endodermal lineages (Yoshida-Koide et al., 2004). Furthermore, the activation of ERK promotes ES cell differentiation and exit from self-renewal (Kunath et al., 2007; Stavridis et al., 2007).

1.3.1.6. Crosstalk between extrinsic factors and the ground state of pluripotency

From the above discussion it is clear that multiple signalling pathways are necessary for the maintenance of ES cells in an undifferentiated state. Regulation of these pathways in different combinations improves the efficiency of ES cell derivation (Ying et al., 2008). For example, inhibition of ERK at the same time as stimulating Wnt signalling enables the derivation ES cells from rat blastocysts (Buehr et al., 2008; Li et al., 2008). However, crosstalk between the different pathways can be instigated by external cues. Several cytokines can activate the same pathways. For example, in addition to the Stat pathway, LIF binding also induces gp130 binding to the SHP2, which results in SHP2 phosphorylation by JAK (Schiemann et al., 1997; Stahl et al., 1995). The phosphorylated SHP2 binds Grb2 and activates the Ras/ERK pathway as well as Gab1 leading to PI3K pathway activation (Takahashi et al., 2005). Moreover, Wu et al (2009) have recently linked LIF and Bmp signalling to SHP2 in ES cells. The authors reported a role for SHP2 in suppressing the Bmp4-SMAD pathway and transcriptional repression of *Id1* and *Id2* (Wu et al., 2009).

Another level at which these pathways interact is at the effector molecules. ERK is directly regulated by Bmps, PI3K, LIF, and Ras signalling pathways (Okita and Yamanaka, 2006; Paling et al., 2004; Qi et al., 2004).

In addition to direct cross talk, different signalling pathways can converge on the same transcription factor network. For example, c-Myc, a key regulator of proliferation and potent inhibitor of ES cell differentiation, is not only activated by LIF through Stat3, but also indirectly by PI3K (Paling et al., 2004) and Bmp signalling via a Ras-ERK activated transcription factors (Qi et al., 2004). Wnt signalling can also directly regulate c-Myc as GSK3-β phosphorylates c-Myc targeting it for proteosome-dependent degradation (Sears et al., 2000).

This cross-talk between signalling pathways can be exploited to derive and maintain ES cells in an undifferentiated state independently of growth factors addition (Ying et al., 2008). LIF and Bmp signals are important to maintain ES cells self-renewing, while FGF-activated ERK induces differentiation. However, ERK remains active in ES cells supplemented by LIF and Bmp. Ying et al (2008) therefore reasoned that LIF and Bmp act downstream of activated ERK to block ES-cell differentiation. Using inhibitors of FGF receptor and ERK (the two inhibitirors are reffered to as 2i), they showed that ES cells remain undifferentiated in the absence of serum or Bmp, but still required LIF for their self-renewal. Without LIF, ES cells degenerated with compromised growth, a phenotype that can be reversed in ES cells by GSK3-β inhibition (3rd inhibitor) (Sato et al., 2004; Ying et al., 2008). The combination of the three inhibitors (the three inhibitors are reffered to as 3i) resulted in a dominant block of commitment and extremely efficient expansion of self-renewing colonies (Ying et al., 2008). Thus, ES cells can be maintained in culture conditions in which pro-differentiation signals have been inhibited and this state has been referred to as the "ground state of pluripotency", implying that the undifferentiated state is the stable default in the absence of developmental inductive signals.

1.3.2. Intrinsic factors

Genetic dissection of factors involved in pluripotency yielded three main transcription factors that maintain pluripotency in the pre-implantation embryo as well as ES cells: Oct4, Nanog and Sox2. However a number of recent studies used

these factors as starting point to dissect the pluripotency network. I will introduce briefly these factors.

1.3.2.1. Oct4

Oct4 was the first pro-pluripotency factor to be identified. It is both conserved in evolution and is common to all pluripotent populations. In chapter three of this thesis I focus on the means by which Oct4 regulates transcription and in chapter 4 on aspects of its coding sequence that are conserved. I briefly review the biology of Oct4 here, but will return to it in greater detail later.

Oct4 is a homeodomain protein encoded by the *Pou5f1* gene. It belongs to the POU class V family of transcription factors. It is expressed in the unfertilised egg and the early embryo before the separation of ICM from trophectoderm in the early blastocyst. Its expression becomes restricted to the ICM and down-regulated in the TE and the primitive endoderm (Pesce and Scholer, 2001). *Oct4* expession persists in the pluripotent epiblast of the pre- and post-implementation embryo, and is down-regulated in an anterior to posterior gradient such that it remains expressed in the region around the primitive streak. Following gastrulation, Oct4 is expressed in the primordial germ cells during their migration and within genital ridges in both sexes (Palmieri et al., 1994). Oct4 is expressed in pluripotent cell lines such as mouse and human embryonic stem cell lines, embryonal carcinoma cell lines (Okamoto et al., 1990; Rosner et al., 1990; Scholer et al., 1989), embryonic germ cell lines, epiblast stem cells, and iPS cells (Do and Scholer, 2009).

Oct4 null embryos develop only to a blastocyst-like stage and the ICM is unable to expand leading to trophectoderm differentiation of ICM cells in blastocyst growth assays (Nichols et al., 1998). However, while blocking trophoblast differentiation is fundamental to pre-implantation development, Oct4 appears to have a multi-faceted role in blocking differentiation during embryogenesis. In vitro analysis of ES cells with tetracycline responsive Oct4 allele established that Oct4 controls the pluripotency of stem cells in a quantitative fashion (Niwa et al., 2000). An increase

of 50% of wild type levels lead to differentiation to endodermal and mesodermal lineages, whereas low levels lead to trophectoderm differentiation (Niwa et al., 2000). Similarly, knock down of *Oct4* in both human and mouse ES cells produced both primitive endoderm and trophoblast differentiation (Hay et al., 2004). Moreover, knock down of *Oct4* homologues in Xenopus result in precocious differentiation of both the prospective neural and marginal cells (Morrison and Brickman, 2006). Oct4, therefore, exhibits an evolutionary role in maintaining pluripotency in the early embryo.

Oct4 is transcription factor that regulates the expression of multiple genes in ES cells that are physiologically relevant for maintaining pluripotency in the ICM such as fibroblast growth factor 4 (*Fgf4*) and zinc finger protein 42 (*Zfp42* or *Rex1*). Oct4 binds the octamer binding sequences within the promoters of target genes, and in cooperation with co-factors such as Sox2 and FoxD3 modulate gene expression either positively or negatively (Pan et al., 2002).

One of the first factors to be shown to bind Oct4 was E1A (Scholer et al., 1991). E1A interacts directly with the basal transcriptional machinery as well as chromatin modifiers such as p300 to modulate the expression of core factors downstream of Oct4 (Zhong and Jin, 2009). Oct4 binds Sox2 to modulate the expression of multiple genes involved in the maintenance of pluripotency such as *Nanog*, *Utf1* and *Fgf4*. Sox2 interacts with Oct4-specific regions outside the POU domain (Ambrosetti et al., 2000). The Oct4/Sox2 protein dimer recruits large multi-protein complexes to control gene expression in ES cells and the early embryo (Rodda et al., 2005). The two proteins interact differentially onto DNA of different target promoter/enhancers (Remenyi et al., 2003). *Fgf4* enhancer contain the POU binding site adjacent to the Sox2 response element, while in the *Utf1* enhancer, the two binding sites are separated by 3bps. This slight difference affects the binding of the two proteins to DNA leading to a varied expression of *Utf1* and *Fgf4* in the early embryo (Remenyi et al., 2003).

Similarly, Oct4 directly binds FoxD3 to block the activation of endodermal specific gene expression of *FoxA1* and *FoxA2* (Guo et al., 2002). FoxD3 activates the

expression of *FoxA1* and FoxA2, but Oct4 binds and inhibits FoxD3 transactivation of the two genes.

1.3.2.2. Nanog

Nanog was identified as a new regulator of self-renewal in ES cells by both functional cDNA expression cloning (Chambers et al., 2003) and digital differential display of expressed sequence tags (ESTs) (Mitsui et al., 2003). Chambers et al (2003) identified Nanog in a screen for LIF-independent ES cell growth, whereas, Mitsui et al (2003) identified Nanog by comparing cDNAs from differentiated and undifferentiated ES cells. Both studies reported that Nanog over-expression compensate for both LIF and Bmp4 signalling. Nanog had been identified previously by Wang et al (2003) as a homeodomain-containing transcript that is predominantly expressed in ICM and they called it ENK (early embryo specific NK) (Wang et al., 2003). *Nanog* expression is initiated in the interior cells of the compacted morula but later confined and highly enriched in ICM although in a more restricted fashion than *Oct4* and its expression becomes down-regulated prior to implantation (Chambers et al., 2003; Wang et al., 2003). *Nanog* is later expressed in the epiblast and becomes restricted to the proximal posterior epiblast (Hart et al., 2004). *Nanog* expression is still detectable in PGCs during migration to and in the genital ridges between E9 and E13 (Chambers et al., 2003; Mitsui et al., 2003). *Nanog* null embryos die at the late blastocyst stage due the absence of the epiblast supporting a pivotal role for Nanog in the maintenance of the epiblast (Chambers et al., 2003; Mitsui et al., 2003). *Nanog* is enriched in human ES and mouse EpiS cells, both of which display epiblast-specific transcriptional signature (Do and Scholer, 2009). Outgrowths of *Nanog-/-* blastocysts give rise to high levels of primitive endoderm differentiation (Chambers et al., 2003; Mitsui et al., 2003).

Nanog role in the maintenance of ES cells remains to be fully appreciated. Its importance in maintaining self-renewal stems from the fact that Nanog overexpression supports LIF-independent self-renewal (Chambers et al., 2003) and its essential role for reprogramming (Silva et al., 2008; Silva et al., 2009). However,

surprisingly, Nanog null ES cells are able to self-renew *in vitro* and express all the common pluripotency markers, as well as contribute to chimera formation (Yates and Chambers, 2005). These cells are identical to wild type ES cells except that they are unable to differentiate into mature PGCs (Yates and Chambers, 2005). Thus, while Nanog is not required for the maintenance of the pluripotent state, it maybe required for its establishment both during development (PGC formation) and in reprogramming.

Nanog is a 305 amino acid phospho-protein that binds DNA as a dimer and deletion of a tryptophan rich interaction motif disrupts dimerisation and sufficient to block LIF-independent self-renewal (Mullin et al., 2008; Wang et al., 2008).

1.3.2.3. Sox2

Sry-related HMG box 2 (Sox2) is a transcription factor belonging to the sequence-specific high mobility group box (HMG) domain-containing protein family related to SRY (Sex-determining Region Y). Sox2 is a 317-amino acid protein containing 3 domains: the N-terminal domain with unknown function, a 79-amino acid DNA HMG binding domain, and a C-terminal transcriptional activation domain (Episkopou, 2005). Like *Oct4* and *Nanog*, mouse *Sox2* expression is first detected at the morula stage and becomes restricted to ICM (Avilion et al., 2003). *Sox2* expression continues in the epiblast until gastrulation when *Sox2* becomes restricted to the presumptive anterior neurectoderm (Avilion et al., 2003). By mid-gestation, *Sox2* becomes expressed in the developing CNS, sensory placodes, in the bronchial arches, and in the gut endoderm (Wood and Episkopou, 1999). *Sox2* null embryos fail to survive after implantation (Avilion et al., 2003). Chimaera analysis demonstrated that there is a cell autonomous requirement for *Sox2* in the epiblast and further in the extra-embryonic ectoderm (Avilion et al., 2003).

Sox2 is expressed in both ES cells and in neural stem cells as well as other cells with restricted developmental potential (Fauquier et al., 2008). In stem cells, Sox2 is essential for self-renewal. Knock down of *Sox2* by siRNA in ES cells promotes differentiation towards trophectoderm as well as other lineages such as mesoderm

and neural lineage (Ivanova et al., 2006).

Sox2 binds DNA cooperatively with Oct4 to mediate the transcriptional regulation of *Nanog* as well as reciprocally regulate each other's expression (Rodda et al., 2005). The interplay between these three factors as well as other key regulators of pluripotency will be discussed in the next section.

The transcriptional networks underlying pluripotency in ES cells mediated by Oct4, Sox2 and Nanog have recently been exploited to reprogram somatic cells back to pluripotent state. Forced expression of Oct4 and Sox2 alongside c-Myc and Klf4 (called OSKM) in mouse and human embryonic fibroblasts was sufficient to generate iPS that are indistinguishable from ES cells (Takahashi et al., 2007a; Takahashi et al., 2007b; Takahashi and Yamanaka, 2006). Nanog appears to be required for this process, but was not part of the reprogramming quartet OSKM (Silva et al., 2008). Nanog also enhances nuclear transfer (Silva et al., 2006).

1.3.3. The core transcriptional network regulating pluripotency

Extrinsic and intrinsic factors maintain ES cell pluripotency by converging on a transcriptional network underpinned by a number of core transcription factors, prominently featuring Oct4, Nanog and Sox2. To begin to understand the mechanisms underlying pluripotency downstream of these three factors, several groups have mapped their target genes in mouse and human ES cells and begun to establish a core network for the regulation of pluripotency.

Boyer et al (2005) used chromatin immuno-precipitation (ChIP) coupled with microarray analysis (ChIP-on-chip) to identify the targets of Oct4, Nanog and Sox2

in human ES cells. Their analysis was limited as the promoters present on the microarray used were pre-selected, and therefore, they were not representative of the whole non-transcribed genome (e.g. long range enhancers were excluded). Also, the authors did not include any functional analysis to verify gene up- or down-regulation. The crucial finding of this study was that Oct4, Nanog and Sox2 co-occupied a significant proportion of target promoters including their own promoters (Boyer et al., 2005). This strongly suggested a feed forward regulatory network as well as a feedback autoregulation loop between the three transcription factors.

Most of the promoters co-bound by three factors control the expression of key developmental regulators encoding transcription factors many of which are homeodomain proteins (Boyer et al., 2005). Some of these co-occupied targets were activated while others were repressed. Activated genes bound by the three factors such as *Zic3* and *Stat3* are mainly involved in proliferation and self-renewal, whereas inactive or repressed genes such as *Pax6* and *Myf5* are involved in differentiation and lineage commitment. This dataset suggested that the feed forward circuitry of pluripotency also involved the repression of lineage specific determinants.

The core circuitry uncovered by Boyer et al (2005) in human ES cells was shown to be conserved in mouse ES cells (Loh et al., 2006; Zhou et al., 2007). Loh et al (2006) mapped Oct4 and Nanog targets only while Zhou et al (2007) extended the bioinformatics analysis to include Sox2. Loh et al used ChIP coupled to paired-end tag sequencing (ChIP-PET) as opposed to promoter arrays. ChIP-PET involves direct sequencing of all immuno-precipitated DNA fragments and is therefore not biased by the selection of tags used to construct the array. This approach is both unbiased and has the advantage of generating targets throughout the genome, rather than just in the immediate upstream of the curated genes (Loh et al., 2006).

In the Loh dataset, a considerable number of genes are co-occupied by Nanog and Oct4. Nanog and Oct4 co-localised to 345 target gene promoters, which represent 44.5% of the total targets compared to 433 genes representing 70% in human ES cells. Using siRNA to knockdown Oct4 and Nanog, Loh et al 2006 reinforced the

existence of an auto-regulatory circuitry and a feed forward network governing mouse ES cell pluripotency. Loss of *Oct4* or *Nanog* reduced the level of their reciprocal expression as well as that of *Sox2*. Expression levels of most genes bound by the two factors changed upon siRNA knockdown of *Oct4* or *Nanog*. Downstream targets involved in self-renewal maintenance such as *Esrrb*, *Rif1*, *Tcf3*, *Jarid2*, *Sall1* and *Rest* were down-regulated, whereas genes involved in lineage commitment such as *Pax6* were upregulated. However, not all bound genes were affected by *Oct4* or *Nanog* loss. This could be explained by functional redundancy in transcriptional regulation or by the possibility of the binding sites being simply non-functional (Loh et al., 2006).

Further analysis of the binding sites of Oct4 and Nanog on DNA revealed that the binding sequences were separated by an average of 25bps on some promoters in the mouse ES cells (Loh et al., 2006). Similar distances are also observed when the binding sites for all three factors were considered in mouse or human ES cells (Boyer et al., 2005; Marson et al., 2008). The binding of these factors close to one another supports the presence of large cooperative protein complexes mediating transcription throughout the network. Such complexes could contribute to the stability of self-renewal, while allowing for a rapid exit from the pluripotent state based on a loss or change in the levels of any one of the three proteins (Chen et al., 2008a).

1.3.3.1 . Hierarchical regulatory network of pluripotency

As discussed above, these global ChIP studies uncovered a hierarchical regulatory network whereby the key three regulators co-occupy the promoters of key developmental transcription factors in mouse and human ES cells. *Pax6*, *Tcl1*, *Stat3*, *Tcf3*, *Esrrb*, *Rif1*, *Sall4* and *Rest* are targets of all three factors in mouse (Loh et al., 2006; Zhou et al., 2007) and human ES cells (Boyer et al., 2005). These developmental factors regulate multiple secondary targets to induce differentiation (e.g. Pax6) or promote self-renewal (e.g. Stat3). The prediction implicit in these datasets is that the core pluripotency factors activate regulators involved in self-

renewal while repressing those involved in lineage commitment. Interestingly a number of these "repressed targets are co-occupied by Polycomb group genes (Boyer et al., 2005; Loh et al., 2006). The implication of these studies is that the repression of lineage determinants by the core pluripotency complex is an essential component of how these proteins maintain ES cell self-renewal. In chapter 3 of this thesis I show that this is most likely not the case and lineage specific repression by Oct4 and associated factors may at best be a secondary failsafe mechanism present in case the stable feed forward pluripotency network collapses.

While the combination of the three core factors appears associated with repressed promoters, these proteins are all representative of generic DNA binding domains, homeodomain, HMG domain and octamer proteins. For this reason, a number of these global ChIP studies may be plagued by artificial binding sites. Moreover, while a number of candidate Oct4 targets came out of the original ChIP studies, direct targets identified in numerous previous studies such as *Fgf4*, *Utf1* and *Pdgfa* were not on the lists of Boyer et al (2005) or Loh et al (2006). A number of functional and in silico approaches to address this problem have been taken. This has focused particularly on Oct4 as it is the best characterized of the three. As mentioned above, genes co-occupied by the core transcription factors are not necessarily real functional targets of these factors. Moreover, even if they are targets, it is not clear what the individual role of components in this large network would be.

Recent studies suggested that a number of regulators to come out of these target screens are essential determinants of pluripotency. An early factor to come out of this analysis is Sall4. Sall4 has been shown to regulate Oct4 expression in ES cells (Zhang et al., 2006), to bind Nanog promoter (Wu et al., 2006), and has recently been shown to regulate distinct circuitries in the two other blastocyst-derived stem cell lines: extraembryonic endoderm (XEN) versus ES cell lines (Lim et al., 2008). In ES cells Sall4 form an extended auto-regulatory network with Oct4, Nanog and Sox2, and its deletion induced differentiation (Lim et al., 2008). While in Xen cells Sall4 regulate Sox and Gata factors to maintain Xen cells self-renewing (Lim et al., 2008). Other targets that were also shown to be required for pluripotency include

Esrrb and Rif1. Knock down of both of these proteins resulted in ES cell differentiation to a flattened fibroblast morphology (Loh et al., 2006).

The chromatin modifiers *Rest* and *Jarid2* are also targets of all three factors that were functionally tested and validated as real functional targets (Loh et al., 2006). Rest was recently shown to control pluripotency in ES and NS cells (Johnson et al., 2008; Singh et al., 2008), while *Jarid1/2c* was shown to modulate H3K9me2 at both *Tcl1* and *Nanog* promoters leading to their activation (Loh et al., 2007)(See Figure1.4).

Other studies tried to identify new genes involved in the maintenance of ES cell self-renewal using functional analysis. Campbell et al (2007) used statistical analysis combined with expression profiling to identify the genes that are activated or repressed by Oct4 as an alternative approach (Campbell et al., 2007). They identified 392 genes that correlated with Oct4 expression and possessed at least one putative composite binding site for Oct4 and Sox2. After validation of 28 of these genes by ChIP and quantitative real time PCR, they confirmed the identification of 26 Oct4 direct transcriptional targets (Campbell et al., 2007). While only 5 genes out of these 26 genes were identified by ChIP-PET and ChIP-chip based studies, 25 genes were also identified by Matoba and colleagues (Matoba et al., 2006). Matoba et al (2006) and Sharov et al (2008) also used a combination of these approaches, to address the issue. By looking at immediate early response to Oct4 depletion using a tetracycline (Tc)-suppressible system, they reached the conclusion that the majority of Oct4 targets were activated by it (Matoba et al., 2006; Sharov et al., 2008).

While all of these studies have examined specific regulators individually, there have also been a number of large scale screens aimed at defining functional determinants for ES cell self-renewal. Ivanova et al (2006) defined a set of 70 genes expressed in ES cells and are rapidly down regulated during differentiation. They stably knocked down all 70 genes by small hairpin RNA (shRNA) and characterized the phenotypes (Ivanova et al., 2006). They identified eight genes that perturb self-renewal upon knockdown. These genes were *Oct4*, *Nanog*, *Sox2*, *Esrrb*, *Tbx3*, *Tcl1* and *Dppa4* as well as the unassigned EST Mm.343880. Knock down of *Oct4*, *Nanog* or *Sox2* resulted in differentiation to multiple lineages. The study showed that *Esrrb* and *Tbx3* inhibit mesodermal, ectodermal and neural crest differentiation, whereas *Tcl1*

loss promotes differentiation towards the neural crest (Ivanova et al., 2006). Interestingly, the down-regulation of *Nanog*, *Sox2*, *Esrrb*, *Tbx3* or *Tcl1* all led to the induction of *Otx2* and *Pitx2* (Ivanova et al., 2006). Otx2 and Pitx2 are key transcriptional regulators of mesodermal and neurectodermal gene expression and are regulated by Oct4 and a Nanog-Sox2 complex respectively (Boyer et al., 2005; Loh et al., 2006; Matoba et al., 2006). Knockdown of any of the eight genes (*Oct4*, *Nanog*, *Sox2*, *Esrrb*, *Tbx3*, *Tcl1*, *Otx2* and *Pitx2*) affected the expression of other pluripotency genes including the core three factors (Ivanova et al., 2006).

1.3.3.2. An extended core transcriptional circuitry

The expansion of the core group of pluripotency regulators as a result of transcriptional profiling, ChIP studies, functional and molecular studies of ES cells has lead to a revision of the target data sets.

Non-biased ChIP-Sequencing has been performed by two different groups: Chen et al (2008) and Kim et al (2008). Chen et al (2008) used ChIP-sequencing to identify targets of 13 transcription factors (TFs) and 2 chromatin modifiers. The aim of their study was to locate the binding sites of the core factors (Nanog, Oct4 and Sox2), the downstream transcription factors of the external signalling pathways (Stat3 and Smad1), the targets of reprogramming factors (c-Myc, n-Myc and Klf4), the downstream targets of other factors implicated in ES cell self-renewal (Zfx, Esrrb and Tcfcp2l1), the cell cycle regulator (E2f1), chromatin insulator TF (CTCF), enhancer-specific acetylatransferase (p300), and PcG methyltransferase (Suz12). The sensitivity of this study appears greater than the previous ChIP studies as they identified new bindings sites for the original core factors.

They also reported that the binding sites formed two major clusters, one constellation of sites that bound Oct4 and the other Myc (Chen et al., 2008b). Sox2, Nanog, Smad1, Stat3, Esrrb, Klf4 and Tcfcp2l1 binding was associated with the Oct4

binding sites, whereas E2f1, Zfx and CTCF were all associated with Myc binding (Chen et al., 2008b). In total, 3583 multiple transcription factor binding loci (MTLs) bound by at least 4 factors were identified. To test the activity of these MTLs, 25 Oct4-MTLs and 8 Myc-MTLs were tested for the capacity to drive heterologous reporter genes. Only the Oct4-MTLs harboured a strong ES cell-specific activity. Interestingly, 87.4% of Smad1 and 56.8% of Stat3 binding sites are within Oct4 centric MTLs (Chen et al., 2008b) supporting the notion that these core MTL sequences integrate signalling with the network. Thus, these Oct4 clusters may represent ES cell specific enhancesomes, which may act as connecting points between the external signalling pathways and the core transcriptional network (Chen et al., 2008b). Further analysis also unveiled an extended feedback loop between the core factors and Stat3 as well as an extended interconnectivity between 11 TFs out of the 13 analysed. The connections between the nodes in the network were categorised into feed forward loop, biparallel motif, fully connected triads and multiple input motifs (Chen et al., 2008a; Chen et al., 2008b).

The Orkin laboratory took a similar approach to mapping the occupancy of nine transcriptional regulators in murine ES cells (Kim et al., 2008). They used *in vivo* biotynilation instead of native antibodies followed by ChIP-seq to analyse the core factors (Oct4, Nanog, Sox2), the remaining reprogramming factors (Klf4 and c-Myc), and Nanog binding proteins identified by Wang et al (2006) (Dax1, Rex1, Zpf281, and Nac1) (Wang et al., 2006). They found that target genes in pluripotent cells fall into two categories: repressed and active genes. Interestingly, the repressed genes tend to be bound by fewer factors, while activated genes were bound by four or more factors (Kim et al., 2008). Also, as shown by Chen et al (2008) and Kidder et al (2008), this study confirmed that Oct4 and c-Myc bound distinct gene promoters and showed similar wiring patterns of gene expression underlying transcriptional network in ES cells (Kidder et al., 2008).

Recently, global ChIP analysis of Tcf3, a downstream component of the Wnt pathway, was performed and compared to the datasets available for the core pluripotency facts (Cole et al., 2008). Tcf3 was found to bind the same genes as the

core factors, and form an autoregulatory circuitry with the core pluripotency factors Oct4, Nanog, and Sox2 (Cole et al., 2008). Tcf3 was shown to be important for maintaining pluripotency and self-renewal both *in vivo* and *in vitro*. It's expression pattern and ablation in pre-implantation embryos support a role in maintaining pluripotency *in vivo* and its loss in ES cells leads to differentiation (Cole et al., 2008).

Moreover, a recent unbiased screen of the entire genome employing gridded shRNA libraries identified 148 genes that were required for the maintenance of ES cells in an undifferentiated state (Hu et al., 2009; Kidder et al., 2008). This extended list included the core factors and some of their direct and indirect targets. However, it also unearthed a novel regulatory network, not evident in the Oct/Sox/Nanog targets. Transcription regulators Cnot3, Trim28, c-Myc and Zfx appear to co-occupy the same targets and are all required for ES cell self-renewal (Hu et al., 2009; Kidder et al., 2008). The targets of this complex are associated with cell cycle, apoptosis and cancer (Hu et al., 2009; Kidder et al., 2008). A similar screen using an Oct4-GFP reporter (Ding et al., 2009) uncovered the core factors and some of their targets, in addition to a new factor, Paf1C (Figure 1.5.). Paf1C regulates the *Oct4* promoter directly and its over-expression blocks ES cell differentiation while its down-regulation leads to differentiation (Ding et al., 2009).

1.3.3.3. MicroRNAs are a part of the hierarchical network

In addition to key signalling pathways, the Oct4, Sox2 and Nanog network contains a number of micro-RNAs (miRNAs) (Loh et al., 2006; Marson et al., 2008). Some of the miRNAs identified were donw-regulated upon differentiation, and their ablation mediated differentiation towards multiple lineages (Marson et al., 2008) while others were upregulated during differentiation (Tay et al., 2008). Interestingly, miRNAs also target Sox2, Nanog and Oct4 and down regulate their expression (Tay et al., 2008). miR-134, miR-296 and miR-470 are induced upon RA differentiation, and they target the 3' UTRs of all three genes (Tay et al., 2008).

Similar results have been obtained in human ES cells (Xu et al., 2009). In human ES cells, miRNA-145 was shown to directly bind the 3'UTRs of *Oct4*, *Klf4* and *Sox2* and down-regulate their expression during differentiation. *miRNA-145* ablation elevates the expression levels of the three factors, and induced differentiation (Xu et al., 2009). Moreover, its expression appears to be regulated by Oct4 (Xu et al., 2009). Together, the miRNAs involved in self-renewal behave with the same dynamics as other factors of the core transcriptional network of pluripotency and exhibit similar feed-back loops.

1.3.3.4. Protein complexes interacting with the core transcription factors

The network so far has been focused on the trinity of Oct4, Nanog and Sox2. However, the notion of co-occupied promoters came out of the ChIP analysis, rather than an unbiased approach to the different complexes containing Oct4, Sox2 and Nanog. To address this, a number of different affinity purification approaches have been used. Using Nanog as a bait to pull down binding proteins, Wang et al (2006) identified a protein interaction network (Wang et al., 2006). Nanog binds directly to ES cell specific transcription factors, Oct4, Dax1, Nac1 and Zfp281 (Wang et al., 2006). All of which were shown later to interact with Nanog but also Oct4 and Sox2 on promoters of target genes (Kim et al., 2008). Nanog also appears to bind a number of global chromatin modifiers such as REST, HDAC2, PcG proteins and components of the SWI/SNF complex (Wang et al., 2006). Members of these complexes are both partners and targets for transcriptional regulation by Oct4, Nanog and Sox2 (Orkin et al., 2008; Wang et al., 2006). Moreover, the components of this network are interdependent, as siRNA knockdown of a single component results in loss of pluripotency. These results suggest the presence of multi-protein complexes that serve as a functional module to maintain ES self-renewal. Using antibodies to the native proteins Liang et al (2008) identified a smaller set of proteins by immuno-precipitation and mass spectroscopy (Liang et al., 2008). This list overlapped

considerably with those derived from BirA tagging (Orkin et al., 2008; Wang et al., 2006) and contained a number of global repressor complexes associated with both Oct4 and Nanog including NurD, Sin3 and Pml complexes. Interestingly, the composition of the newly identified NuRD complex was distinct from known NuRD-associated complexes. The NuRD complex associated with Oct4 and Nanog called NODE (for Nanog and Oct4 associated deacetylase) contains preferentially Mta1 while Mbd3 and Rbbp7 are either lacking or present at low levels (Liang et al., 2008). Knockdown of *Mta1*, *Mta2* or *Hdac2* lead to impairment of ES self-renewal. In particular, *Mta1* loss upregulated the expression of endoderm lineage-specific genes *Gata6* and *FoxA2* recapitulating *Nanog* knockdown (Liang et al., 2008).

1.3.4. Epigenetic regulation of pluripotency

The central role of transcription factors in pluripotency brings an intriguing possibility that chromatin and DNA modifications (i.e. epigenetics), the major substrates of transcription, are important for pluripotency. Epigenetic modifications modulate transcriptional regulation in specific genomic regions and underpin the cross talk between transcription factors and basal transcriptional machinery (e.g. TATA binding and RNA II [i.e PolII]). For example, DNA methylation is indicative of silenced loci, trimethylation of lysine 9 and lysine 27 of histone 3 (H3K9me3 and H3K27me3, respectively) indicates repressed genes, whereas trimethylation of lysine 4 of histone 3 (H3K4me3) and acetylation of histone 3 or 4 tails represent active genes. Epigenetics is known to be important for oocyte to zygote transition, and determines the totipotentiality of the zygote (Surani et al., 2007). Pluripotent cells are epigenetically different from either the zygote or more differentiated cells even though they have the same genetic makeup. Therefore, epigenetics is one of the major determinants of lineage commitment and differentiation.

1.3.4.1. Chromatin and chromatin modifications of ES cells

Studying chromatin structure of ES cells and mapping different chromatin modifications across the ES cell genome suggested that chromatin is more

permissive in self-renewing ES cells and becomes restrictive upon differentiation (Surani et al., 2007).

Meshorer and colleagues assessed the global chromatin status of ES cells by measuring the exchange rate of chromatin associated protein, they used fluorescent recovery after photo-bleaching (FRAP) to analyse the mobility of chromatin proteins in ES cells versus differentiated cells. They reported that major chromatin architectural proteins such as the linker histone H1, histones H2B, H3, and heterochromatin-associated protein (HP1α) bind loosely and briefly to the chromatin of self-renewing mouse ES cells making it more accessible to transcription factors and chromatin modifiers (Meshorer and Misteli, 2006). However, upon differentiation, the dynamic nature of these architectural proteins decreases. Moreover, they also showed that heterochromatic markers are dispersed in mouse ES cells and become more concentrated in distinct foci in differentiated cells (Meshorer and Misteli, 2006).

Other studies also investigated the global chromatin state of ES cells using different assays. Azuara and colleagues used replication time as an indicator of chromatin state (Azuara et al., 2006). Early replication during S phase is a characteristic of transcriptionally active regions of the genome, and therefore is an indicator of open chromatin. However, late replication is an indicator of heterochromatin, and gene repression (Azuara et al., 2006). They reported that lineage specific genes that are not active in ES cells replicted earlier in pluripotent cells than differentiated cells, and had both active and repressive chromatin marks, suggesting that these lineage specific genes are poised for expression in ES cells and held in check until their appropriate time of expression (Azuara et al., 2006). This in turn suggests that the chromatin of pluripotent cells is more permissive than that of differentiated cells.

Global analysis of histone modifications also suggests that there is an overall loss of the open chromatin state with differentiation. ES cells are characterised by high levels of activating H3K4me3, that is lost and replaced by repressing H3K27me3 and H3K9me3 upon differentiation (Bernstein et al., 2006; Meshorer and Misteli, 2006; Mikkelsen et al., 2007; Reik, 2007). Similarly ES cells exhibit high levels of histone acetylation which decrease with lineage commitment (Meshorer and Misteli, 2006).

1.3.4.2. Bivalent domains and poised expression in ES cells

As discussed in earlier sections, the core pluripotency factors were found to activate genes involved in self-renewal while repress the ones implicated in lineage commitment (Boyer et al., 2005; Loh et al., 2006). In fact, it was reported that most of the co-occupied targets are repressed rather than activated (Boyer et al., 2005; Loh et al., 2006). While this has yet to be formally proven, a number of groups have looked at the chromatin states at some of these genes. ChIP mapping of H3K4me3 and H3K27me3 revealed that a number of these genes contain both active and repressed chromatin marks (Azuara et al., 2006; Bernstein et al., 2006). While this was initially puzzling, the presence of these marks or "bivalent domains," was thought to represent a class of genes repressed, but poised to be activated. Interestingly genome-wide ChIP-seq (or chip-chip) for these modifications suggest they are mostly associated with differentiation-specific genes that are inactive or expressed at very low levels in self-renewing ES cells (Azuara et al., 2006; Bernstein et al., 2006). It was suggested that bivalent domains offer a model for transcriptional regulation of tissue-specific genes (Bernstein et al., 2007). They are silenced in ES cells but are activated upon differentiation (Mikkelsen et al., 2007). However, this model is challenged by findings that OCT4, NANOG, and SOX2 are marked by only active chromatin mark (H3K4me3) in undifferentiated human ES cells, but acquire a bivalent domain configuration upon differentiation (Pan et al., 2007).

Many studies question the functional relevance of these bivalent domains in ES cells. They report that bivalent domains are not unique to pluripotent cells, but were also observed in progenitor cells such as neural progenitor cells, committed cells such as MEFs and terminally differentiated cells such as lung fibroblasts and T cells (Barski et al., 2007; Mikkelsen et al., 2007; Pan et al., 2007).

1.3.4.3. Polycomb group proteins

The H3K27 trimethylation in the "bivalent domain," is catalysed by the polycomb repressor complex 2 (PRC2) and this modification is generally associated with PRC2

targets. PRC2 consists of three proteins including proteins: embryonic ectoderm development (Eed), suppressor of zest 12 (Suz12) and the H3K27 methyltransferase: enhancer of zest homologue 2 (Ezh2) (Pietersen and van Lohuizen, 2008). Interestingly about half of the genes that exhibited H3K27me3 were also bound by at least one of the pluripotency transcription factors Oct4, Nanog, or Sox2 (Boyer et al., 2005; Boyer et al., 2006b). However, ES cells mutant for specific PRC components do not display a defect in pluripotent gene expression, but rather have strong phenotypes of differentiation (Boyer et al., 2006a). Null embryos for PRC components are embryonic lethal and exhibit gastrulation defects, suggesting a failure in lineage specification (Faust et al., 1998; O'Carroll et al., 2001; Pasini et al., 2007; Pasini et al., 2004; Pasini et al., 2008b). Specific defects in individual PRC components appear to have similar phenotypes in ES cells. *Eed* null cells exhibit global reduction of H3K27me3 and propensity to differentiate but retain the expression of pluripotent factors (Boyer et al., 2006b). Similarly, *Suz12-/-* ES cells self-renew but are differentiation-defective (Pasini et al., 2007; Pasini et al., 2004). Moreover, Chamberlain et al (2008) have shown that PRC2 is dispensible for ES cell maintenance in pluripotency state, but is required for differentiation of ES cells (Chamberlain et al., 2008).

One of the functions of PRC2 is to recruit PRC1. PRC1 core components are Ring1A, Ring1B and Bmi1. Embryos mutant for the PRC1 components also exhibit early embryonic lethality and gastrulation defects (Pasini et al., 2008a). ES cells that are null for both Ring1A and Ring1B loose the typical morphology of ES cells, and have an impaired proliferation (Endoh et al., 2008). ChIP-on-chip against specific PRC components Phc1 and Rnf2 (PRC1) as well as Suz12 and Eed (PRC2) identified 512 targets bound by the complete set of these proteins and exhibited the repressive mark H3K27me3 (Ku et al., 2008). The dataset was enriched in developmental regulators including the *Hox*, *Dlx*, *Irx*, *Lhx*, *Pou*, *Pax* and *Six* gene families. These genes were all upregulated in PRC knock downs suggesting a direct role for PcG proteins to silence these targets in ES cells. Similarly, during ES cell differentiation these genes loose H3K27me3 (Ku et al., 2008; Lee et al., 2007a). Lee et al (2006) reported similar findings using the PRC2 component *SUZ12* in human ES cells (Lee et al., 2006).

These finding support a role of PRC proteins in suppressing the expression of key lineage determinants in ES cells, but this does not fit with either their embryonic or ES cell phenotypes, which suggests that these proteins play a more fundamental role in differentiation.

1.3.5. Cross-talk between epigenetic and the core factors of pluripotency

A number of candidate Oct4, Sox2 and Nanog targets are chromatin modifiers. For example Oct4 represses genes such as *Set7*, the methyltransferase catalyzing mono-methylation of histone H4 lysine 20 (H4K20) and *Hdac6* (Babaie et al., 2007). Oct4 also regulates *Jmjd1a* and *Jmjd2c*, genes encoding the H3K9me2 demethylases (Figure1.4) (Loh et al., 2007). Oct4 also interacts with Stat3 to upregulate the expression of the PRC2 core protein *Eed* (Ura et al., 2008). Moreover, *Oct4* down-regulation leads to inactivation of other PRC proteins such as *Suz12* and *Phc1* (Endoh et al., 2008). Oct4, Nanog and Sox2 target a number of other chromatin factors including *Eset*, *Myst3* and *Smarcad1* (Boyer et al., 2005). As a result, Oct4 indirectly regulates histone modification and the epigenetic status of an array of non-target genes through such chromatin modifiers. Moreover, pluripotency factors have also been shown to interact directly with chromatin modifiers. As discussed above, Oct4, Nanog and Sox2 co-occupy many of the targets of PRC2 to mediate gene silencing (Lee et al., 2007a). Furthermore, Ring1b, a PRC1 component, has also been shown to be recruited to its targets in an Oct4-dependent manner to repress gene expression (Endoh et al., 2008). Nanog and Oct4 also interact with the histone deacetylase complex NuRD (P66b and HDAC2), polycomb group (YY1, Rnf2, Rybp) and SW1/SNF chromatin remodelling complex (Wang et al., 2006).

Some of the pluripotency factors are themselves also regulated by epigenetic factors. The *Oct4* locus is progressively repressed during ES cell neural differentiation. This repression is accompanied by the loss of active epigenetic marks and their replacement with repressive marks (Aoto et al., 2006). Additionally, germ cell nuclear factor (GCNF) interacts with Dnmt3 and recruits methyl-binding domain

proteins (MBDs) to the *Oct4* locus to mediate its silencing by DNA methylation (Sato et al., 2006).

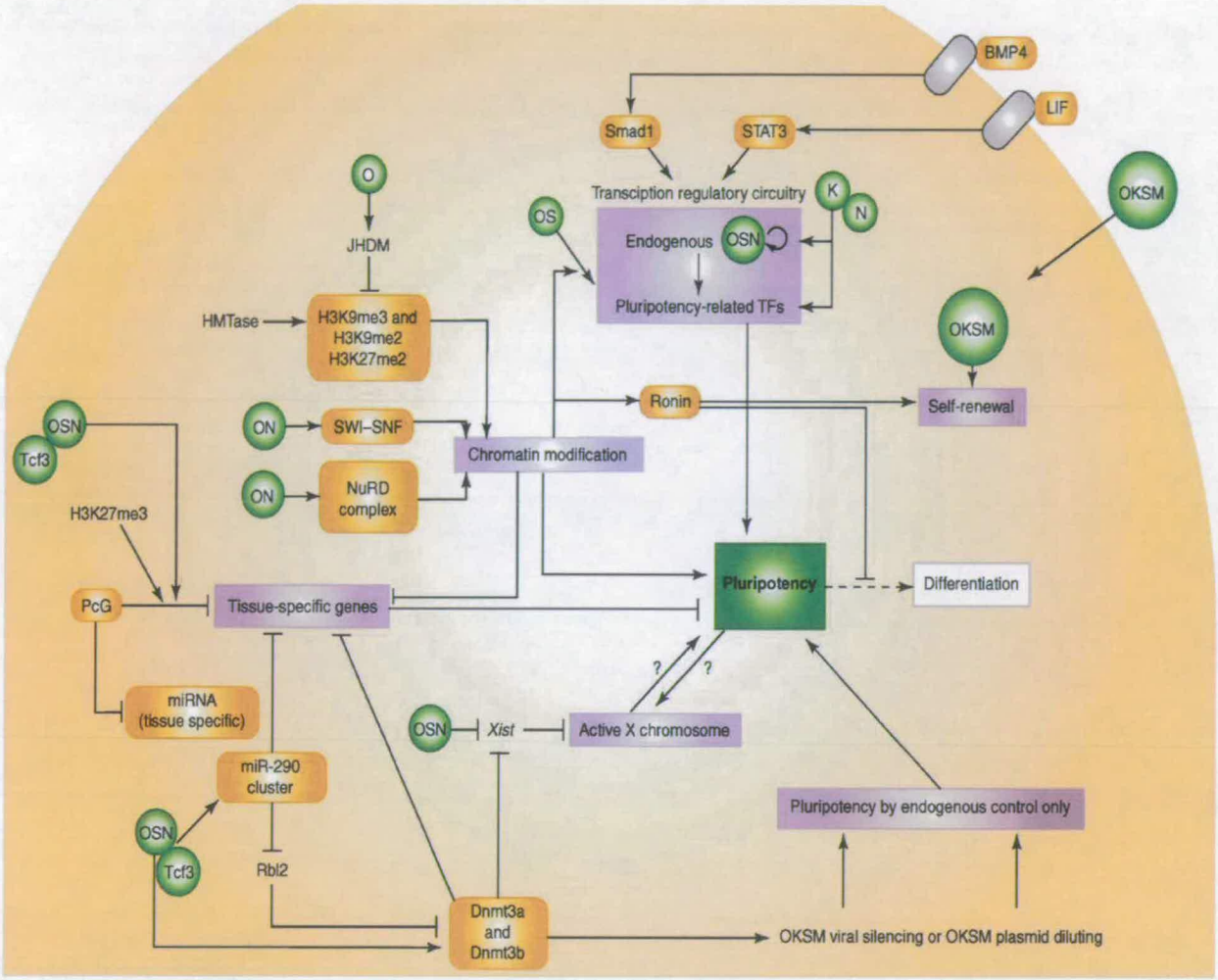


Figure 1.4- Genetic and epigenetic mechanisms of pluripotency

The pluripotent state is maintained by a crosstalk between 1) factors of transcriptional network, 2) external pathways, and 3) epigenetic determinants. The pluripotency is induced by the transduced transcription factors (OSKM) during the process of reprogramming. OSKM maintain a transcriptional network akin to the core ES cell transcriptional network (OSN). OSN cooperate with DNA and chromatin modifying enzymes as well as miRNAs to maintain pluripotency *in vitro* (ES cells) and *in vivo* (ICM).

The abbreviations are as follows: JHDM, histone demethylases of the JmjC family; K, Klf4; M, c-Myc; miRNA, microRNA; N, Nanog; O, Oct4; PcG, polycomb group; S, Sox2; TFs, transcription factors; Xist, X inactivation specific transcript. '?' not yet illustrated.

Adopted from (Do and Scholer, 2009).

Dnmt3a and Dnmt3b are not only bound to the *Oct4* locus but have been recently shown to mediate also *Nanog* silencing after ES cell differentiation (Figure1.5) (Li et al., 2007). Loh et al (2007) further showed that Oct4-mediated activation of *Jmijd1a* and *Jmjd2* positively regulate pluripotency-associated genes. While Jmijd1a demethylates H3K9me2 at the promoters of *Tcl1*, *Tcfcf2l1*, and *Zfp57*; Jmjd2c demethylates H3K9Me3 at the *Nanog* promoter leading to its activation (Loh et al., 2007).

1.4. Detail analysis of Oct4

1.4.1. Embryonic expression of Oct4

Pou5f1 is a maternally expressed gene and the Oct4 protein is present at low levels in the oocyte until fertilisation (Palmieri et al., 1994). The expression becomes weak but detectable up until the 8 cell stage where the levels of Oct4 transcript and protein increase to the point that all cells in the morula stage are Oct4-positive (Palmieri et al., 1994). Zygotic expression of Oct4 begins at the four- to eight-cell stage. At these early stages, uniform expression of Oct4 is observed in all blastomeres. During compaction, the outermost cells down-regulate Oct4 expression and differentiate towards the TE lineage. This is mediated by a negative feedback loop between Oct4 and Cdx2 (Niwa et al., 2005). The expression of *Oct4* is restricted to the inner cell mass from the compacted morula stage onward. The primitive endoderm cells that start to differentiate and migrate alongside the inner surface of the TE transiently express higher levels of Oct4 protein than in the ICM, something that was not detected by RNA localization studies (Palmieri et al., 1994; Pan et al., 2002). Post-implantation, Oct4 expression is further restricted to the epiblast, and at gastrulation this expression becomes further restricted in a posterior-proximal direction where Oct4 remains expressed in the primitive streak region until the end of gastrulation (Boiani and Scholer, 2005). Following gastrulation Oct4 becomes restricted to the primordial germ cells (Boiani and Scholer, 2005). The expression of Oct4 strongly correlates with cells that maintain the widest range of potency at each stage of the embryo development. Moreover, most multi-lineage potent stem cell lines (ES, EG,

EpiS and iPS) express Oct4 (Do and Scholer, 2009). Upon the differentiation of all these types of embryonic pluripotent cells, Oct4 decreases confirming that Oct4 is marker of pluripotency. Several studies have reported that Oct4 is expressed in somatic stem cells. However, it has been recently shown that results can be explained by an expression of a pseudogene (Lengner et al., 2008).

1.4.2. Regulation of Oct4 expression

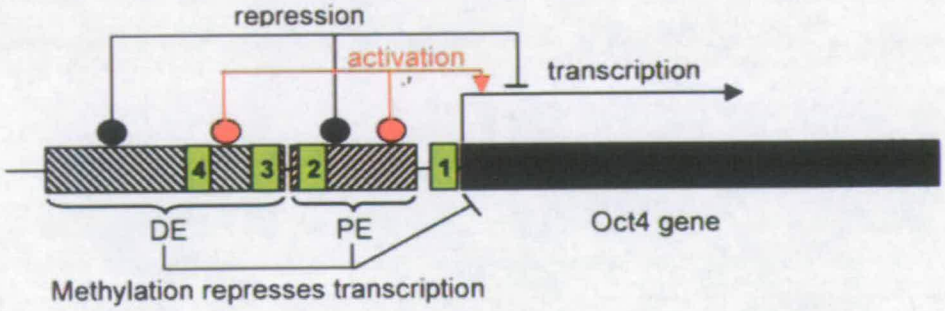
Based on experiments in transgenic lines with reporter genes, 18 Kbps of the promoter that recapitulates accurately the endogenous expression of *Oct4* was characterised (Yeom et al., 1996). The TATA-less promoter of *Oct4* has been shown to contain three regions essential for its restricted expression before and after implantation. The proximal promoter (PP, see Figure5a) is within the first 250 bp of the transcriptional start site, it contains three overlapping half-sites of hormone response element (HRE), which is known as canonical retinoic response elements (RARE) (Schoorlemmer et al., 1994; Sylvester and Scholer, 1994). The proximal promoter of *Oct4* also contains an overlapping GC box (5'-(G/T)GGGCGG(G/A)(G/A)(C/T)-3') that is bound by the zinc finger transcription factor Sp1 (Minucci et al., 1996). However, *Sp1* null mice have normal expression of Oct4, suggesting that other related zinc finger proteins compensate for its function in development such as the Kruppel-like factor (KLF) protein. Klf2, 4 and 5 are essential for the maintenance of ES cells and good candidate regulators for this region (Jiang et al., 2008).

Adjacent to the proximal promoter are the proximal and distal enhancers (PE and DE, respectively see Figure5a), both of which are essential for stage and cell type-specific expression. DE drives expression in the ICM and ES cells, while the PE regulates Oct4 expression in the ectodermal lineage as well as P19 EC cells (Okazawa et al., 1991; Yeom et al., 1996). Using *in vitro* DNA footprinting, the precise binding sites of protein complexes on PE and DE were identified. Two GC-box rich sites 1A and 2A were found in PE and DE respectively, both are crucial for driving expression from the two enhancers but showed no ES specific expression (Minucci et al., 1996; Pesce and Scholer, 2001).

However the factors involved in regulating *Oct4* promoter activity are not well undersood. Heterodimers of nuclear hormone receptors RAR and RXR were shown to positively regulate Oct4 expression in undifferentiated P19 cells, through binding to RARE sequence in the PP. However, during RA induced differentiation of ES cells, this activation can be blocked by the binding of the orphan nuclear receptors COUP-TF1, COUP-TF11(ARP1) and EAR2 to the RARE sequence as inactive homo- or hetero-dimers with RXR (Ben-Shushan et al., 1995; Sylvester and Scholer, 1994). Unlike the orphan nuclear receptors stated above, Liver Receptor Homologue-1 (LRH-1) is required for the maintenance of *Oct4* expression at the epiblast stage (Gu et al., 2005). Moreover, LRH-1 was shown to bind to PE and DE and activates *Oct4* expression in undifferentiated ES cell (Gu et al., 2005). The other known positive regulators of Oct4 expression include steroid factor 1(SF1) (Barnea and Bergman, 2000), GA repeat binding protein α (GABP-α) (Kinoshita et al., 2007), and the orphan nuclear receptor Tr2, which also acts as a repressor upon SUMOylation (Park et al., 2007). Furthermore, Oc4 positively regulates its own expression through cooperative binding with Sox2 to DE (Chew et al., 2005; Okumura-Nakanishi et al., 2005). Essentially, Sox2 and Oct4 co-occupy a new site 2B, located 30bps downstream from site 2A within DE, in both ES cells and embryos (see Figure 5a). This new site was necessary and sufficient in combination with site 2A to drive DE expression in reporter assays (Chew et al., 2005; Okumura-Nakanishi et al., 2005).

Oct4 is repressed in differentiation and is silenced in somatic tissues. This silencing has been shown to be mediated through the recruitment of DNA methyltransferases and chromatin modifying enzymes to the promoter and enhancers of Oct4 (Ben-Shushan et al., 1993). The orphan receptor, GCNF, has been shown to interact with PP and to recruit Dnmt3, suggesting its involvement in the epigenetic silencing of *Oct4* during gastrulation (Fuhrmann et al., 2001; Sato et al., 2006). G9a, a SET-containing H3K9me histone methyltransferase, was also shown to mediate *Oct4* inactivation during early embryogenesis (Feldman et al., 2006).

(a)



(b)

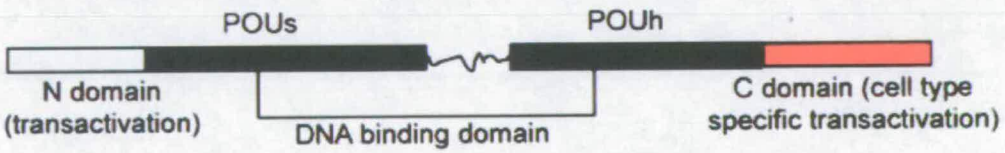


Figure 1.5- The structure of the *Oct4* gene and protein

(a) The structure of the upstream regulatory elements of the *Oct4* gene. There are 4 regions that are highly conserved among human, bovine and mouse *Oct4* promoter/enhancer elements (boxes 1 through 4). Conserved region 1 (CR1) is downstream of PE and immediately upstream of exon 1. Each enhancer contains multiple potential binding sites for transcription factors that can either induce activation (red) or repression (black) of *Oct4* expression. Methylation in CRs represses Oct4 expression in differentiated and somatic cells. Abbreviation: DE, distal enhancer, and PE, proximal enahncer

(b) Schematic illustration of Oct4 domains. POUs and POUh are important for DNA binding, while the C and N termini are involved in binding to co-activators/co-repressors. Abbreviations: N-: -amino, C-: Carboxyl-, POUs: POU-specific domain, POUh: POU-homeo domain.

Adopted from (Pan et al., 2002).

Transcription factors were also shown to be involved in the negative regulation of *Oct4*. Tcf3 has been recently shown to repress *Oct4* expression in embryogenesis. It binds within the PP of *Oct4* and represses gene expression through the recruitment of co-repressors C-terminal binding protein (CtBP), Split2 and Groucho (Tam et al., 2008). Similar to Tcf3, the homeodomain transcription factor Cdx2 was also reported to repress *Oct4* gene expression. Niwa et al (2005) proposed that reciprocal inhibition of Oct4 and Cdx2 controls the first lineage decision in mammalian development, the formation of the extra-embryonic trophectoderm (TE) from the ICM (Niwa et al., 2005).

1.4.3. Structure of Oct4

Pou5f1 is a highly conserved gene in vertebrates including fish and amphibians that is expressed only in early development within the pluripotent cells. In mouse, the gene comprises of five exons located on chromosome 17 within the t-region. Genes within the t-region are associated with lethality. The protein encoded by *Pou5f1* gene, Oct4, also known as Oct3/4, is 352 amino acid and belongs to the POU domain family of transcription factors.

The POU domain family of transcription factors was defined following the observation that the mammalian proteins Pit-1, Oct-1, and Oct2 and the *Caenorhabditis elegans* protein unc-86 shared a region of homology, known as the POU domain (Clerc et al., 1988; Finney et al., 1988; Herr et al., 1988; Robertson, 1988; Sturm et al., 1988; Sturm and Herr, 1988). The POU domain is a unique DNA binding domain that is composed of two highly conserved regions, the POU-specific and the POU-homeodomain, joined by a variable linker (Figure 1.5b) (Botfield et al., 1992; Verrijzer et al., 1992). POU proteins have been identified from various species, and are grouped into seven classes based on the amino acid sequence of their POU domains and their linker region (Hinkley et al., 1992; Spaniol et al., 1996; Wegner et al., 1993).

Most POU proteins bind to an eight base pair regulatory DNA element known as the octamer motif ATGCAAAT (Scholer, 1991). Despite the fact that the POU-specific

and POU-homeo domains are both independently able to bind to DNA in sequence specific manner (Aurora and Herr, 1992; Klemm and Pabo, 1996), POU proteins require both domains for high affinity binding to their DNA targets (Aurora and Herr, 1992). This suggests that the linker region acts to increase the local concentration of DNA binding domains and therefore increasing the site-specific DNA binding (Ryan and Rosenfeld, 1997).

The crystal structure of the Oct1 and Pit1 POU domains bound to their DNA recognition sites as a monomer and homodimer, respectively, revealed the three dimensional structure of the two sub-domains as well as the adaptability of POU proteins (Jacobson et al., 1997; Klemm et al., 1994). The POU-specific domain consists of four α helices grouped around a hydrophobic core. The helices form a helix-turn helix (HTH) motif (Klemm et al., 1994). The POU-homeodomain also forms a HTH motif as it consists of 3 α helices, of which the first two are relatively variable between different POU proteins, however the third is highly conserved (Jacobson et al., 1997; Klemm et al., 1994).

Oct-1 and Pit-1 crystal structures also revealed the disordered structure of the linker region, which suggests that its main function is simply to connect the POU-specific and POU homeodomains, and reinforce its role in increasing DNA binding specificity (Ryan and Rosenfeld, 1997). In addition to this role, the linker region might contribute to the flexibility of POU domain in recognizing various DNA elements by enabling the POU-specific domain to take various orientations relative to the POU-homeodomain (Ryan and Rosenfeld, 1997). POU proteins were also shown to have different types of trans-activation domains located outside of the POU domain, depending on amino acid composition or acidity (Mitchell and Tjian, 1989). The structure of Oct4 protein and the role of its sub-domains and their relevance to ES cell self-renewal will be discussed in details in the last result chapter of this thesis.

1.4.4. Oct4 homologues

The expression of *Oct4* in gastrulation suggests that it has a fundamental role in the lineage specification that should be conserved in evolution. This notion was supported by the identification of *Oct4* homologues in *Xenopus laevis* that are able to support murine ES cells in the absence of *Oct4* (Morrison and Brickman, 2006). Knocking down of Oct4 homologues in Xenopus leads to precocious differentiation in multiple lineages indicating that Oct4 may have a conserved role in blocking differentiation in vertebrates. Interestingly, murine Oct4 can rescue aspects of the Xenopus knock down phenotype (Morrison and Brickman, 2006).

Xenopus laevis and *Xenopus tropicalis* both contain three Oct4 homologues: *Xlpou60*, *Xlpou25* and *Xlpou91* organised in tandem within a region of the genome that is analogous to the mouse Oct4 t-locus (Morrison and Brickman, 2006). Their expression recapitulates the expression of Oct4 in mouse pre-implantation development. *Xlpou60* is maternally expressed and decreases by the late blastula stage, while *Xlpou91* and *Xlpou25* are zygotically expressed and continue to be expressed alongside Xlpou61 during gastrulation in cells that did not undergo involution (Morrison and Brickman, 2006).

The expression patterns and roles of the Xenopus homologues suggest that these members of the PouV protein family have roles in both neural development and isthamic specification (Snir et al., 2006). The evolutionary conserved role of PouV homologues was also highlighted by the characterisation of a Zebrafish PouV protein, Pou2/Spg (Takeda et al., 1994). Spg is a mutation in Pou2 that affects mid hind brain specification and isthamic development (Takeda et al., 1994). Pou2 is located in a chromosomal region syntenic to the Xenopus region (Belting et al., 2001; Hauptmann et al., 2002; Lunde et al., 2004). Pre-gastrulation expression of Pou2 recapitulates the expression of mouse Oct4 both in the oocyte and zygote (Foygel et al., 2008; Howley and Ho, 2000) and its loss in the early stages of development leads to comparable phenotypes (Foygel et al., 2008; Lunde et al., 2004; Reim et al., 2004). Furthermore, the proximal promoter of *Pou2* harbours octamer binding sequences that regulate *Pou2* expression in a similar fashion to Oct4 auto-regulatory loop (Parvin et al., 2008). However, there is some controversy about

whether Pou2 is a true orthologue as the role of Pou2 in the isthmus appears conserved with Xenopus but not with mouse (Takeda et al., 1994). Moreover, Pou2 also has a role in endoderm induction in cooperation with a Sox protein known as Cassanova that does not have a mammalian homologue (Foygel et al., 2008; Lunde et al., 2004; Reim et al., 2004). As *Pou2* is also not an effective substitute for *Oct4* in ES cells, it is possible that teleosts have diverged significantly in their use of the Oct4 network (Morrison and Brickman, 2006).

Other studies have uncovered conserved Oct4-related genes in different mammalian species including Rhesus monkeys, cows and pigs. These genes show similar pre-implantation expression pattern to *Oct4* (Kirchhof et al., 2000). Furthermore, Niwa and colleagues reported the conservation of *Oct4* in the platypus that faithfully recapitulates *Oct4* expression in embryogenesis (Niwa et al., 2008). Although it was reported that chicken genome does not contain an Oct4 homologue (Soodeen-Karamath and Gibbins, 2001). Lavial et al (2007) have identified a chicken *Oct4* (*cOct4*) using subtractive hybridization of cDNAs from chicken ES cells and EBs (Lavial et al., 2007). *cOct4* and chicken *Nanog* homologue cooperate to maintain chicken ES cells in an undifferentiated state, showing the conservation of Oct4 mediated mechanisms underlying pluripotency. Like the Xenopus protein, Xlpou91, cOct4 is able to rescue self-renewal in *Oct4* null mouse ES cells (Lavial et al., 2007).

1.4.5. Activation and repression functions of Oct4

Oct4 protein is a transcription factor that was reported to bind regulatory regions of many genes including the *Rex-1* gene (Ben-Shushan et al., 1998) Rosfijord and Rizzino 1994), the *Fgf4* gene (Schoorlemmer and Kruijer 1991; (Yuan et al., 1995), and the human platelet-derived growth subunits of human chorionic gonadotrophin gene (*hCG*) (Liu et al., 1997; Liu and Roberts, 1996).

In *vivo*, Oct4 can act as a repressor or an activator of target genes (Ben-Shushan et al., 1998). It can directly activate or repress transcription of target genes when the *cis*-binding site is positioned proximal to the promoter (Scholer et al., 1991). One example of Oct4 direct activation is the PDGFα receptor (*PDGFαR*) gene. Oct4 was

shown to bind to one of two alternative promoters of the PDGFα receptor (PDGFαR) gene to activate transcription. In fact, it was reported that mutating the octamer motif present in this promoter decreased its activity, whereas overexpressing Oct4 in differentiated Tera2 cells specifically enhances it (Harry J. Kraft, 1996).

In contrast, Oct4 can also act as a direct repressor. For instance, its binding to the enhancer of the immunoglobulin heavy chain gene was reported to repress transcription in F9 EC cells. Lenardo and colleagues reported that a 300-base pair fragment of the heavy chain enhancer was inactive in F9 EC cells and that mutating the octamer motif present within the enhancer increased its activity in F9 EC cells, suggesting that Oct4 is directly repressing the immunoglobulin heavy chain gene in pluripotent cells (Lenardo et al., 1989). Another example of direct gene repression by Oct4 is the repression of *hCG* genes, which are required for implantation and the maintenance of pregnancy (Liu et al., 1997; Liu and Roberts, 1996). It was reported that Oct4 binding sites were found within the proximal promoters of these genes, and that mutating the octamer binding site in the *hCG-β* gene promoter abolished its Oct4-mediated repression. Moreover, mouse Oct4 binding to these regulatory elements in co-transfection assays was shown to repress transcription from both human *hCG α* and *β* gene promoters in JAr choriocarcinoma cells (Liu et al., 1997; Liu and Roberts, 1996).

Oct4 also acts indirectly by binding to other cofactors. Oct4 can indirectly repress *FoxA1* and *FoxA2,* members of the Forkhead Box (Fox) family, without binding to their promoters. Guo and colleagues reported that in transient cotransfection assays Oct4 inhibited FoxD3 activation of the *FoxA1* and *FoxA2* promoters. As Guo et al showed that Oct4 interact with FoxD3 protein both *in vitro* and after cotransfection in 293 cells, they proposed that this repression is mediated by Oct4 interaction with the DNA binding domain of FoxD3 (Guo et al., 2002). Similarly, Oct4 can also mediate the silencing of tau interferon genes (*IFNτ*) by binding to their activator (Ets-2) DNA binding domain, and quenching its transactivation function (Ezashi et al., 2001). Ezashi and colleagues reported that the upregulation of the *IFNτ* resulting from the overexpression of *Ets-2* in human JAr choriocarcinoma cells is inhibited by coexpression of Oct4. Moreover, the fact that expression from luciferase reporters driven by mutant *IFNτ* promoters lacking potential octamer sites was still silenced by

Oct4 suggested that Oct4 mediates the repression via protein-protein interaction, especially that Oct4 and Ets-2 were coimmunoprecipitated after being expressed together in Jar cells. This was further endorsed by the ability of Oct4 and Ets-2 to form a complex in vitro in the absence of DNA through binding of the POU domain of Oct4 to DNA binding domain of Ets-2.

When acting over a long distance to activate target genes, in a similar fashion to repression, Oct4 requires cofactors that can serve as a bridging factor between Oct4 and the basal transcription machinery (Scholer et al., 1991). The adenovirus (Ad) E1A oncoprotein was reported to mimic the function of such cofactors in pluripotent cells. Another oncoprotein, HPV-E7, was also reported to have a similar role in Oct4-mediated gene activation (Brehm et al., 1999; Brehm et al., 1998; Yeom et al., 1996).

Oct4 can also synergize with other transcription factors bound to a close *cis*-acting element of target genes. It binds cooperatively with the high mobility group (HMG) transcription factor Sox2 within the 3'-untranslated region (UTR) of the *Fgf4* gene to activate transcription in a stem cell specific manner (Yuan et al., 1995). Another example of this Oct4/Sox2 synergistic activity is the activation of the stem cell specific gene *Utf1* (Nishimoto et al., 1999).

Figure 1.6 summarises Oct4 modes of actions on different target genes.

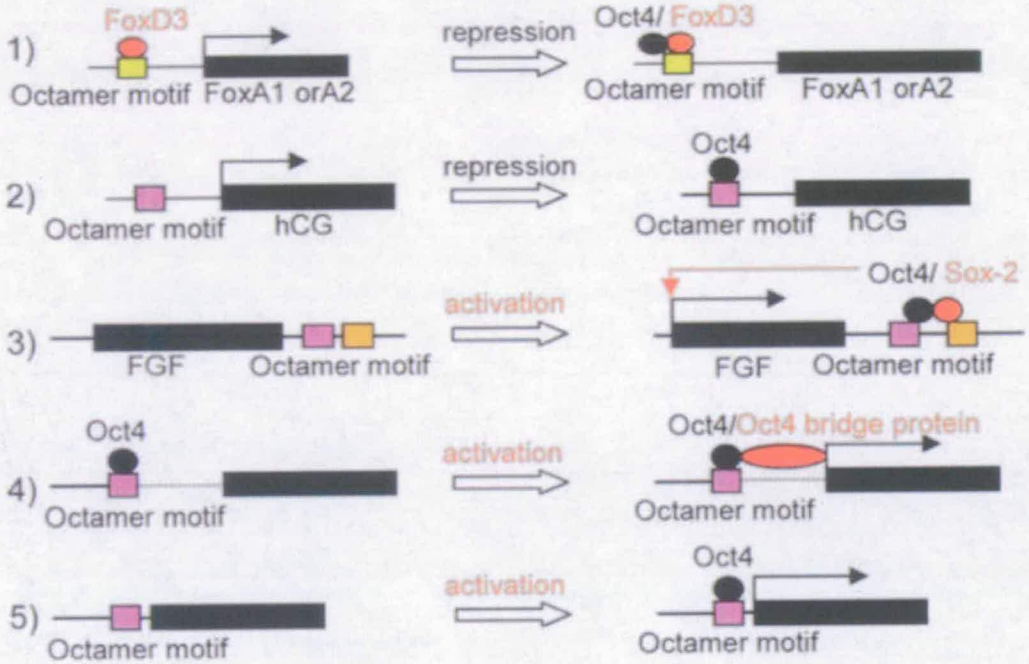


Figure 1.6- Activation and repression by Oct4 on different target genes

1) Oct4 represses FoxA1 and FoxA2 gene expression indirectly by interacting with the DNA binding domain of their activator FoxD3 and neutralizing its function.
2) Oct4 represses the human chorionic gonadotropin (*hCG*) genes directly by binding to their promoters.
3) Oct4 can synergize with other factors like Sox2 to activate transcription from the Fgf4 gene.
4) When octamer site is located further away from the target gene promoter, Oct4 requires adaptor proteins to bridge it to the basic transcription machinery for transcriptional activation.
5) Oct4 can activate directly the PDGF α receptor (PDGFαR) gene by binding to octamer sites.

Adopted from (Pan et al., 2002).

Aims

In chapter 3 we test whether lineage repression is an essential component of the ability of Oct4 to regulate differentiation in embryonic populations.

In chapter 4 we set to identify PouV protein domains that are necessary for ES cell self-renewal.

CHAPTER 2

Materials & Methods

2. Materials and Methods

Unless otherwise stated, chemicals were ordered from Thermo-Fisher, oligonucleotides were obtained from Sigma-Genosys. The water used of MilliQ (18.2mΩ) quality.

2.1. Mammalian cell culture

2.1.1. Non-ES cell lines

The human embryonic kidney cell line HEK293 and the human choriocarcinoma cell line JAR were cultured in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen) supplemented with 10% Foetal calf serum (Invitrogen) and penicillin-streptomycin (Roche). The cells were passaged every 2 to 3 days. Penicillin-streptomycin was not added to the medium during DNA transfections as the antibiotics interfere with lipofectamine and hinder cell growth post-transfection.

2.1.2. ES Cell lines

E14Tg2a and ZHBTc4 ES cell lines were grown in ES medium (see below) on gelatin-coated (0.1%) plates or flasks (Iwaki) at 37^0c in a 7% CO_2 incubator.

2.1.2.1. ES medium

500ml	Glasgow Minimum Essential Medium (GMEM) (Sigma-Aldrich)
11ml	Sodium pyruvate/L-Glutamine (5.5 ml of 100mM sodium pyruvate +
5.5 ml	200mM L-Glutamine)
51ml	Foetal calf serum (FCS) (Invitrogen)
5.5ml	100x non-essential amino acids (Invitrogen)
555µl	0.1M 2-mercaptoethanol.
555µl	LIF (prepared by ISCR tissue culture staff)

2.1.3.2. Trypsinisation of ES cells

ES cells were washed with PBS after the medium was aspirated. 0.025% trypsin/PBS (v/v) was added to the cells, followed by incubation for few minutes at 37^0c. The cells are collected and washed with PBS by centrifugation at 250g for 3 mins.

2.1.3.3. Passage of cells

After trypsinisation, cells were resuspended in fresh pre-warmed ES medium. 1:5 of the cells were plated onto gelatinised flasks that were kept at 37^0C for few minutes. The cell lines were passaged every 3 to 4 days with media changes every 48hours.

2.1.2.4. Freezing ES cells

Around $1x10^6$ cells were collected and washed once in the freezing mix (10% DMSO in ES cell medium). The pellet was resuspended in 1ml freezing mix and transferred to a cryotube vial (Nunc) to be placed in the -80^0C freezer. The next day the cryo-preserved cells were transferred to the liquid N_2 storage tank (-170^0C).

2.1.3.5. Thawing mouse ES cells

The vial was removed from liquid N_2 and thawed immediately in a 37^0C water bath. The thawed cells were transferred to a universal tube containing 9ml pre-warmed ES cell medium. The cells were pelleted by centrifugation at 200g followed by a PBS wash.

2.2. Transfection of DNA into cells

2.2.1. Transient transfection of ES, HEK293 and JAR cells

2.2.1.1. Luciferase reporter assays

The transfection of cells was performed using Lipofactamine 2000 (Invitrogen) as per manufacturer's instructions. Twenty-four to forty-eight hours post transfection, the cells were collected, lysed and monitored for both firefly and renilla luciferase activity using the dual luciferase reporter assay system (Promega) according to manufacturer's instructions.

2.2.1.1.1. Testing the transcriptional activity of activator and repressor forms of Oct4 and Xlpou91

ZHBTc4 ES cells ($1x10^5$) were plated on a 24-well plate with 2μg/ml tetracycline (Tc). Twenty-four hours later, 75 ng of reporter plasmids and 150 ng of the test plasmid were transfected according to (Brickman et al., 2001). An internal reference plasmid consisting of the SV40 promoter driving Renilla luciferase was used with all transfections (Promega).

2.2.1.1.1a. Reporter plasmids

Reporter plasmids are a kind gift from Prof. Hitoshi Niwa (Riken Institute, Japan) and were constructed by introduction of the regulatory elements listed below into the pGL3-Basic vector (Promega) (Niwa et al., 2002).

- **6x Octamer-binding motif luciferase reporter:** 6W enhancer containing 6 copies of oligonucleotides with octamer-binding motif from the mouse immunoglobulin heavy chain gene enhancer upstream of the *tk* promoter.
- ***Fgf4*-enhancer luciferase reporter:** 460-bp PCR-amplified genomic DNA carrying the *Fgf4* enhancer placed upstream of the tk-luc cassette in plasmid ptk-luc.

- **pTK luciferase reporter:** 168-bp *Bam*HI-*Xho*I fragment derived from pBLCAT2 containing herpes simplex virus thymidine kinase (*tk*) promoter.

2.2.1.1.1b. Test constructs

The expression vector used was the pCAG-IP plasmid, a kind gift from Prof Hitoshi Niwa (RIKEN institute, Japan). pCAG-IP plasmid was constructed as fellows: "the pacbGHpA cassette from pPGKpurobGHpA was amplified by PCR with oligonucleotide primers 5'-AAGCTTATCATGACCGAGTACAAGC-3' and 5"-GAGCCCCTGCAGGTTCTTTCCGCC-3" using *Pfu* polymerase (Promega.). The product was digested with *Bsp*HI and *Pst*I and ligated to the *Pst*I-*Not*I backbone and the *Not*I-*Nco*I internal ribosome entry site (IRES) fragments of pCAG-IZ to give pCAG-IP. cDNAs listed below were introduced between the *Xho*I and *Not*I sites of pCAG-IP" (Niwa et al., 2002).

- Empty vector (pCAG-IP with no insert)
- Xlpou91
- Oct4
- Xlpou91λVP2 (Brickman laboratory reagent)
- Oct4λVP2 (Brickman laboratory reagent)
- Xlpou91λEnR (Brickman laboratory reagent)
- Oct4λEnR (Brickman laboratory reagent)
- DNA binding-mutant Oct4
- DNA binding-mutant Oct4λVP2
- DNA binding-mutant Oct4λEnR

Primers to generate the cDNAs are listed in the "Cloning primers" section.

2.2.1.1.2. Testing the transcriptional of Oct4 activator form on the β-subunit of human chorionic gonadotropin (hCG-β) promoter

JAR cells were transfected at 60% confluency on a 6-well plate. 2μg of luciferase reporter constructs with 2μg of test constructs using Lipofectamine 2000 (Invitrogen) as per tmanufacturer's protocol. pRL-SV40 (Promega) was used as internal control.

2.2.1.1.2a. Reporter construct

hCG-β luciferase reporter: -325 hCG-β-luc construct was a kind gift from Prof. Michael Roberts (University of Missouri, USA) (Liu and Roberts, 1996).

2.2.1.1.2b. Test construct

cDNAs listed below were introduced between the *Xho*I and *Not*I sites of pCAG-IP (Niwa et al., 2002).

- Empty vector (pCAG-IP with no insert)
- Oct4
- Oct4λVP2

2.2.1.2. Transient transfection of activator and repressor forms of Oct4 for protein analysis

HEK293 cells plated onto 6 well plates were tranfected with 2μg of vector DNA using lipofectamine 2000 (Invitrogen) following the manufacturer's protocol. Twenty four hours post transfection, cells were lysed to be used for western blotting analysis.

2.2.2. Transfection of ES cells for stable integration

On the day of transfection $2x10^7$ ZHBTc4 cells were collected and washed twice in PBS and resuspended in 800 μl of PBS. 100μg of linearized DNA was added to the cells and the mixture was transferred to an electroporation cuvette (BioRad) and left at RT for 3mins. Electroporation was carried out with a BioRad GenePulser using the following settings: capacitance, 3μF; Voltage, 800V; and time, 0.1 secs (constant). After electroporation, the cells were transferred into pre-warmed ES cell media containing 2μg/ml of tetracycline (Tc) before plating at a density of $2x10^6$ cells per pre-gelatinised 10cm dish.

2.2.2.1. Vectors used for stable transfections

The vectors used to derive cell lines from the ZHBTc4 parental cell line, were constructed on the pCAG-IP backbone (Niwa et al., 2002), and were as follows:

- Oct4
- Oct4λVP2
- Oct4λEnR
- DNA binding-mutant Oct4
- DNA binding-mutant Oct4λVP2
- DNA binding-mutant Oct4λEnR
- Xlpou91
- Xlpou91λVP2
- Xlpou91λEnR
- Xlpou25
- N91P25C25
- N25P91C91
- N25P25C91
- N91P91C25
- N25P91C25
- N91P25C91
- 3x Flag Oct4
- 3x Flag Xlpou91
- 3x Flag Xlpou25
- 3x Flag N91P25C25
- 3x Flag N25P91C91
- 3x Flag N25P25C91
- 3x Flag N91P91C25
- 3x Flag N25P91C25
- 3x Flag N91P25C91
- Pou2
- N91P2C2
- N2P91C91

- N2P2C91
- N91P91C2
- N2P91C2
- N91P2C91

See "Cloning primers" section for the primers used to generate the cDNAs.

2.2.2.2. Puromycin selection

48 hours post transfection, medium was replaced with one containing 2μg/ml of puromycin (Sigma) to start selection for stable transfectants. The medium was changed every two days and the selection was kept for 9 days.

2.2.2.3. Picking ES cell colonies

After 9 days of antibiotic selection, the medium was removed and cells were washed twice with pre-warmed PBS. 5μl trypsin was aspirated in a yellow tip and expelled over a colony. Using the tip, the colony was scraped and pulled up and transferred onto a gelatinised well of a 96 well plate. 180μl of ES cell medium with 2μg/ml puromycin was added following by pipetting up and down to break up the colony. The colony was expanded to be used in downstream experiments. Each expanded colony was frozen.

2.2.2.4. List of cell lines derived

- Oct4
- Oct4λVP2
- Xlpou91
- Xlpou91λVP2
- 3x Flag Oct4
- 3x Flag Xlpou91
- 3x Flag Xlpou25
- 3x Flag N91P25C25

- 3x Flag N25P91C91
- 3x Flag N25P25C91
- 3x Flag N91P91C25
- 3x Flag N25P91C25
- 3x Flag N91P25C91

2.3. Cellular analysis

2.3.1. Staining of ES cells

2.3.1.1. Alkaline phosphatase staining

The cells were washed once with pre-warmed PBS and fixed for 45 seconds. The cells were washed once with dH_2O. Four mls of stain solution was added to 10 cm^2 plate and incubated at RT for 25mns in the dark. The stain was removed and washed once with dH_2O. The plates were air dried before examination under the microscope.

2.3.1.1a. Fixative Solution

The fixative solution was stored at 4^0C and prepared as follows:

- 25mls citrate solution (18mM citric acid, 9mM sodium citrate, 12mM NaCl)
- 8ml Formaldehyde
- 65ml acetone

2.3.1.1b. Stain solution

An alkaline phosphatase staining kit (sigma) was used according to the manufacturer's protocol. Briefly, 400µl FRV alkaline solution was mixed with 400µl sodium nitrate solution and incubated for 2mins at RT. The alkaline/nitrite mixture was added to 18mls of dH_2O. Finally, 400µl Napthol was added to the mix.

2.3.2. MTS proliferation assay

To determine the proliferative potential of different ES cell clones, the CellTiter 96 Non-Radioactive Cell Proliferation Assay (Promega) was used. The assay is a colorimetric alternative to [3H]thymidine incorporation assays designed to determine cell proliferation and cytotoxicity that uses the tetrazolium reagent MTS (3-(4,5-dimethylthiazol-2-yl)-5(3-carboxymethonyphenol)-2-(4-sulfophenyl)-2H

tetrazolium). The Promega assay was carried according to the manufacturer's protocol. Briefly, ES cells were plated into the wells of a 96 well plate. Following this, an MTS assay was performed by adding 20 µl of CellTiter 96 AQ_{ueous} One Solution Reagent to each well. The plate was then incubated for 1 h at 37^0C followed by absorbency reading at 490 nm.

2.3.3. Differentiation of ES cells

2.3.3.1. LIF withdrawal

For differentiation experiments, cells were seeded in monolayers on gelatin-coated dishes in GMEM medium with 10% FCS and 2µg/ml of tetracycline (Tc) and puromycin but without LIF. The day of seeding was defined as day 0 of differentiation. Medium was replaced every two days and cells were harvested at the indicated time points for downstream experiments. As a control, all cell lines used were also grown in the presence of LIF.

2.3.3.2. Oct4 overexpression

Cells were seeded in monolayers on gelatin-coated dishes in GMEM medium with 10%FCS, 2µg of puromycin, and the presence of LIF. Oct4 overexpression condition is achieved by removal of tetracycline (Tc) from the medium. The day of seeding was defined as day 0 of differentiation. Medium was replaced every two days and cells were harvested at the indicated time points for downstream experiments. As a control, all cell lines used were also grown in the presence of Tc.

2.3.3.3. Neural differentiation

Cells were plated on gelatinised plates at density of 10^4 cells/cm^2 in N2B27 medium (Stem cell sciences inc) for 7 days. Medium was changed every day. As negative control, cells were plated on gelatinised plates in ES cell medium at the same density. Cells were collected at day 7 or 8 for downstream experiments.

2.3.4. FACS analysis

Fluorescence-activated cell sorting (FACS) was used to analyse the percentage of ES cell clones expressing the cell surface markers SSEA1, E-Cadherin and PECAM1 (for the antibodies see table 2.1.). Cells were collected into the cell dissociation buffer (Invitrogen) and incubated at 37^0C for 10 minutes. Single cell suspension was achieved by pipetting throμugh a syringe needle. Following PBS washes, cells were resuspended in 500 μl FACS buffer (1xPBS and 10% FCS) at a concentration of 10^6 cells/ml. To mark apoptotic cells, 5μl/10^6 cells of 7AAD solution (BD Pharmingen) was added to cells. Primary antibodies raised against cell surface markers were added to the resuspended cells at a dilution of 1:1000 and incubated on ice for 10mins followed by three washes in FACS buffer. Cells were resuspended in FACS buffer with the secondary antibody diluted at 1:1000 and washed as before. Finally, the cells were resuspended in 500μl FACS buffer and analysed in a FACS calibur flow cytometer (BD Biosciences). Data were analysed using the Cell Quest software (BD Biosciences).

2.3.5. Immunofluorescence

Cells grown in 12 well plates were washed twice in PBS followed by fixation in 4% paraformaldehyde. Cells were permeabilised in PBST (1xPBS, 0.1% Triton X100). Blocking was performed by adding 1% bovine serum albumin (BSA) to PBST to the cells and left for 30 mins at RT. Primary antibodies (See Table 2.1 for details) incubated overnight at 4^0C followed by three washes in PBST for 10 minutes each. Alexa Fluor-conjugated secondary antibodies (See Table 2.1 for details) were diluted in block solution and added to the cells. The cells were incubated at RT for 1 hour with the DAPI solution (1:1000). Finally, cells were washed 3 times in PBST, and either analysed under the microscope directly or stored in PBS.

TABLE 2.1 Antibodies used for Western blotting

	Protocol / Dilution			
	WB	**IF**	**FACS**	**Supplier/Reference**
Oct3\4	1 in 1000	1 in 1000		Santa Cruz (sc-9081)
Nanog		1 in 1000		Chambers *et al* 2003
Sox2		1 in 1000		Boyer *et al* 2005
Flag	1 in 1000			Sigma-Aldrich (F3165)
VP16	1 in 1000			Sigma-Aldrich (V4388)
Engrailed	1 in 1000			AbCam (ab12454)
β-Actin	1 in 10000			SigmaAldrich (A3853)
β-tubulin	1 in 1000			Santa Cruz Biotechnolgy (sc-58667)
E-cadherin			1 in 1000	Santa Cruz Biotechnolgy (sc-52328)
PECAM1			1 in 1000	Santa Cruz Biotechnolgy (sc-28188)
SSEA1			1 in 1000	Santa Cruz Biotechnolgy (sc-21702)
Anti-Rabbit-HRP (2^{nd})	1 in 5000			GE Healthcare (NA931)
Anti-Mouse-HRP (2^{nd})	1 in 5000			GE Healthcare (MA934)
Anti-Rabbit- Alexa Fluor (2^{nd})		1 in 1000		Invitrogen (Alexa fluor-488)
Anti-Mouse- Alexa Fluor (2^{nd})		1 in 1000		Invitrogen (Alexa fluor-546)

2.3.6. Xenopus experiments

2.3.6.1. RNA used for embryo injection

To generate RNA for injections, pCS2+ plasmids harbouring the target cDNAs *Oct4λVP2, Xlpou91λVP2, Oct4λEnR, Xlpou91λEnR, Oct4, Xlpou91,* DNA binding mutant-*Oct4* and *Bmp4* were linearised with *BssHII* and used as a template for RNA synthesis using SP6 polymerase (Promega).

2.3.6.2. RNA injection in Xenopus embryos

Xenopus laevis embryos were obtained by *in vitro* fertilisation cultured as described before (Slack et al., 1984). Embryos were injected at the two cell stage into both blastomeres with RNA to a final concentration of 500pg per embryo or morpholino oligonucleotides. Embryos were staged according to (Nieuwkoop, 1997) and collected at stages 9 and 10 for RNA extraction and at stage 10 for *in situ* hybridisation.

2.3.6.3 *In situ* hybridisations

For *in situ* hybridisations analysis, embryos were treated following the protocol of (Gammill and Sive, 2001). *Eomes, Xvent2, Bmp4, Fgf8b* and *Gsc* antisense riboprobes were synthesised from linearised plasmids using T7 polymerase (Promega) following standard protocol using digoxigenin labelled nucleotide (Roche).

2.3.6.4. Morpholinos

Antisense morpholino oligonucleotides (Genetools) specific for *Xlpou25*, *60* and *91* were previously described (Cao et al., 2006; Morrison and Brickman, 2006) and injected to a final concentration of 40ng per embryo (80ng for *Xlpou91*). Control morpholinos were injected to a final concentration of 160ng per embryo.

2.3.6.5. RNA Isolation and real-time RT-PCR

Total RNA was prepared from 5 embryos using RNeasy mini kit (Qiagen) as per the manufacturer's protocol for animal cells with DNase1 treatment. RNA (500ng) was used as a template for cDNA synthesis using Superscript III (Qiagen). Real time PCR was carried out using a Lightcycler 480 (Roche) with LightCycler FastStart DNA Master^{PLUS} SYBR Green 1 (Roche). Primers for PCR were designed either

using Primer 3 (Rozen and Skaletsky 2000) as described previously (Cao *et al* 2006; Morrison and Brickman, 2006) (See Table 2.2 for primer sequences and details). Standard curves were generated from plasmids produced by TOPO-TA cloning. Samples were normalised to ornithine decarboxylase (*Odc)*.

2.4. DNA analysis

2.4.1. DNA quantification

The concentrations of DNA preparations were determined by measuring the optical density of the samples at 260nm (OD_{260}) using Nanodrop (ThermoScientific). DNA concentration was calculated using the relationship: 1 OD unit at 260nm = 50µg/ml DNA.

2.4.2. DNA agarose gel electrophoresis

DNA fragments were loaded with 6xDNA loading buffer (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 30% (v/v) glycerol) and separated by electrophoresis on 1% (w/v) agarose gels, prepared in 1x TAE (40mM Tris-acetate and 2mM EDTA) and 1µg/ml ethidium bromide.DNA fragments were run at a constant voltage of 80V for 1h in 1x TAE. Samples were loaded alongside double stranded DNA molecular weight markers (Promega). DNA fragments were visualized on a dual intensity ultraviolet transilluminator (UVP).

2.4.3. Extraction of DNA from agarose gels

DNA restriction fragments separated on agarose gels were purified using the QIAquick Gel Extraction Kit (Qiagen), according to the manufacturer's instructions. After agarose separation, DNA was viewed with a long wavelength ultraviolet emission (UV) transilluminator and the required band was cut out using a scalpel and transferred to an Eppendorf tube. The gel slice was dissolved by adding 3 volumes (per weight of gel slice) of QG buffer (QIAquick spin kit, Qiagen) and incubated at 50°C for 10 min. The sample was transferred to a QIAquick spin column and centrifuged for 1 min after which the DNA in the column was washed with 0.75ml of PE buffer (QIAquick spin kit). Traces of wash buffer were removed by centrifugation for 1 min. Finally, the DNA was eluted by adding 30µl of ddH_2O and centrifugation for 1 min at full speed in a microcentrifuge. The DNA was collected in a 1.5ml microfuge tube.

2.4.4. Bacterial strains

In most cases, the XL1-Blue Sub-cloning-Grade competent *E. coli* cells (Stratagene) were used for plasmid preparation. This strain has tetracycline resistance and its genotype is: *recA1 endA1 gyrA96 thi-1* h*sdR17supE44relA1 lacz*. The XL1-Gold Ultra competent cells (Stratagene) were used for manipulation of plasmid DNA. XL10-Gold strain genotype is : Tet_r *Δ(mcrA)183 Δ(mcrCBhsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F' *proABlacI^{q}ZΔM15 Tn10* (Tet_r) Amy Cam_r]$_a$ .The XL1-Blue supercompetent cells (Stratagene) were used for preparation of plasmids manipulated for mutagenesis (as advised by the manufacturer). The XL1-Blue supercompetent strain genotype is: Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173end A1supE44thi-1 recA1 gyrA96 relA1 lac.

2.4.5. Media and maintenance

Competent cells were grown in either Luria-Bertani (LB) broth or LB agar. LB broth was made up by mixing 20g/l of LB broth powder (Sigma-Aldrich) with ddH_2O and autoclaved for 20 min at 120°C. LB agar was made by mixing 37g/l of LB agar pellets with ddH_2O following by 20 mins autoclaving. Where necessary, ampicillin (Sigma-Aldrich) was added to 100μg/ml final concentration.

2.4.6. Bacterial transformations

A single 50μl aliquot of frozen competent bacteria was thawed on ice. One μl of plasmid DNA (0.5-1μg/μl) was mixed with the thawed cells and left on ice for 30 min. Cells were subjected to heat shock by immersing them in a 42°C water bath for 45s and subsequently kept on ice for a further 2 min. One ml of LB medium (without ampicillin) was added and the transformed bacteria were incubated at 37°C in a bacterial shaker for 1h. Transformed bacteria were plated onto agar plates containing ampicillin and incubated overnight at 37 °C.

2.4.7. Cloning

2.4.7.1. Restriction endonuclease digestion

Restriction digestions of DNA plasmids were performed using restriction enzymes from Roche. The 10x buffer provided for each enzyme was diluted to 1x in the final reaction mixture. Incubation time and temperature were adjusted according to the conditions specified by the manufacturer.

2.4.7.2. Ligation

For this, the digested DNA was first extracted from the gel, purified and eluted in 10µl of ddH_2O. Ligation reactions were subsequently set up using 1µl T4-Ligase (New England Biolabs), in 1 xT4 Ligation buffer in a total volume of 15µl and suspended in water at 4°C overnight. Samples were left at RT for at least 2 h prior to transforming with competent *E.coli* cells. To scan for correct ligation products quick protocols such as restriction digestion followed by sequencing.

2.4.7.3. Polymerase chain reaction

PCR reactions were carried out using the Qiagen Taq DNA polymerase kit (Qiagen). PCR reactions were set up in a total volumes of either 50µl or 20µl consisting of 1x PCR buffer (Qiagen), 1x Q solution (Qiagen), 0.2nM dNTP (Qiagen), 0.6µM of each forward and reverse primers, 0.5µl of Taq polymerase (Qiagen) and water. The PCR was carried out using a Gene Amp® PCR system 9700 (Applied Biosystems) using the following program:
94°C for 5 min, 5 cycles of denaturing at 94°C (30 s), annealing at 58 °C (30 s, the temperature was changed for each primer) and extending at 72 °C (3 min) followed by 21-25 cycles of the same but annealing temperature of 55 °C (30 s) and finally a further extension of 72°C for 7 min.

2.4.7.4. TOPO cloning of PCR products

To clone PCR fragments, the TOPO-TA cloning kit (Invitrogen) was used as per the manufacturer's instructions.

2.4.7.5. Site directed mutagenesis

Mutagenesis of Oct4 to generate Oct4 DNA-binding mutant was carried out using the QuikChange® Site-Directed Mutagenesis Kit (Stratagene. The primers used for creating point mutations are:

Forward: 5'-TGGGCTAGAGAAGGATGTGCCCCGGGTATGGTTCTGTAACCGGC-3'

Reverse: 5'-GCCGGTTACAGAACCATACCCGGGGCACATCCTTCTCTAGCCCA-3'

Briefly, sample reactions were prepared in a total volume of 50µl consisting of ddH_2O, 5µl of 10 x Reaction Buffer (Stratagene), 50ng of double stranded plasmid DNA, 125ng of each forward and reverse primers, 1µl of 25mM dNTP (Stratagene) and 1µl of *PfuTurbo* DNA polymerase (2.5U/µl). Mutagenesis was performed using a temperature cycler to first denature the plasmid and to anneal the oligonucleotide primers containing the desired mutation. The Thermo-cycler programme was adjusted as below: 95°C for 30 s, 12 (for point mutations) or 18 cycles (for deletions) of denaturing at 95°C (30 s), annealing at 55°C (1 min) and 68 °C (6 min). Following temperature cycling, the reaction tubes were left on ice for 2 min to cool the reaction to below 37°C. XL1-Blue (Stratagene) supercompetent cells were transformed as previously described. DNA was extracted from transformed bacteria using the QIAprep Spin Miniprep kit (Qiagen) and was sent for sequencing to ensure that the desired mutation has been achieved prior to large scale amplification of the mutated plasmids.

2.4.7.6. DNA sequencing

Each sequencing reaction was prepared using 3.2pmol of either the forward primer or reverse primer targeted to the plasmid backbone (Sequencing services at Edinburgh University), 1µg DNA in a total volume of 10µl made up with ddH_2O. Prepared samples

were sent to the sequencing service at Edinburgh University (Ashworth building). Sequence analysis was performed using DNA STAR Lasergene® v7.0 software (DNASTAR).

2.4.7.7. Cloning primers

1. DEF-mutant Oct4
   - Forward: 5'-CTGTATCCGCAGCAGCACCCCCAGGTCCCCACTTTGGCA-3'
   - Reverse: 5'-TGCCAAAGTGGGGACCTGGGGGTGCTGCTGCGGATACAG-3'
2. DEF-mutant Xlpou91
   - Forward: 5'-CAAGGTCCTGCAGCAGCACCCCAAGTGATGCCCTC-3'
   - Reverse: 5'-GAGGGCATCACTTGGGGTGCTGCTGCAGGACCTTG-3'
3. Generation of Flag-tagged constructs
   - 3x Flag Forward: 5'-CCGGTCGACATGGACTACAAAGAC-3'
   - 3x Flag Reverse with A: 5'-CCGCTCGAGACTTGTCATCGTCATC-3'
   - 3x Flag Reverse: 5'-CCGCTCGAGCTTGTCATCGTCATC-3'
4. Generation of N-terminus-deleted Oct4
   - ΔNOct4 Forward: 5'-CCGCTCGAGCCGCCATGAAGGAGCTAGAACAG-3'
   - Oct4 Reverse: 5'-ATAGTTTAGCGGCCGCTCAGTTTGAATGCATGGGAGA-3'
5. Generation of N-terminus-deleted Xlpou91
   - ΔNXlpou91 Forward: 5'-CCGCTCGAGCCGCCATGGGGGAGATGGAGC-3'
   - Xlpou91 Reverse: 5'-ATAGTTTAGCGGCCGCTCAGTTGCCTTGTTTACCCA-3'
6. Generation of C-terminus-deleted Oct4
   - Oct4 Forward: 5'-CCGCTCGAGATGGCTGGAC-3'
   - ΔCOct4 Reverse: 5'-ATAGTTTAGCGGCCGCTCACTCAATACTTGATCT-3'
7. Generation of C-terminus-deleted Xlpou91
   - Xlpou91 Forward: 5'-CCGCTCGAGCCGCCATGTATAACCAACAGACCTACC-3'
   - ΔCXlpou91 Reverse: 5'-ATAGTTTAGCGGCCGCTCAGGGGTACACCTGGCGC-3'
8. Generation of chimeric proteins between Xlpou91 and Xlpou25
   - Xlpou91 Forward: 5'-CCGCTCGAGCCGCCATGTATAACCAACAGACCTACC-3'
   - Xlpou91 Reverse: 5'-ATAGTTTAGCGGCCGCTCAGTTGCCTTGTTTACCCA-3'
   - Xlpou25 Forward: 5'-CCGCTCGAGCCGCCATGTACAGCCAACAGCCCTTCC-3'
   - Xlpou25 Reverse: 5'-ATAGTTTAGCGGCCGCTCAGCCAATGTGGCCCCCCA-3'
   - 91/25 3'end Nbr1:
     5'-GCAAACTGCTCCATTTCTGAAGAATTAGGGGCTTCCTCCTC-3'

- 91/25 5'end Nbr2:
  5'-GAGGAGGAAGCCCCTAATTCTTCAGAAATGGAGCAGTTTGC-3'
- 25/91 3'end Nbr3:
  5'-GCAAACTGCTCCATCTCCCATTCGCTGGGAACCTCCTCCTC-3'
- 25/91 5'end Nbr4:
  5'-GAGGAGGAGGTTCCCAGCGAAGGGGAGATGGAGCAGTTTGC-3'
- N-terminus + homeo 91 Nbr1:
  5'-GCCGTCATTCTCCTCAACGGTGGGGTACACCTGGCGCTTGCC-3'
- C-terminus 91 Nbr2:
  5'-GGCAAGCGTCAGGGAATGCCCTATATTAGGGAGAATGGCGGG-3'
- N-terminus + homeo 25 Nbr3:
  5'-CCCGCCATTCTCCCTAATATAGGGCATTCCCTGACGCTTGCC-3'
- C-terminus 25 Nbr4:
  5'-GGCAAGCGCCAGGTGTACCCCACCGTTGAGGAGAATGACGGC-3'

9. Generation of chimeric proteins between Xlpou91 and Pou2
   - Xlpou91 Forward: 5'-CCGCTCGAGCCGCCATGTATAACCAACAGACCTACC-3'
   - Xlpou91 Reverse: 5'-ATAGTTTAGCGGCCGCTCAGTTGCCTTGTTTACCCA-3'
   - Pou2 Forward: 5'-CCGCTCGAGATGACGGAGAGAG-3'
   - Pou2 Reverse: 5'-ATAGTTTAGCGGCCGCTTAGCT-3'
   - Reverse primer A: 5'-GCTCCAAATCTTCAGAATTAGGGG-3'
   - Sense primer B: 5'-GCCCCTAATTCTGAAGATTTGGAG-3'
   - Sense primer C: 5'-ACTCTGACTACTGGGGAGATGGAG-3'
   - Reverse primer D: 5'-CTCCATCTCCCCAGTAGTCAGAGT-3'
   - Sense primer E: 5'-CTAGCTTTGCCCTATATTAGGGAG-3'
   - Reverse primer F: 5'-CTCCCTAATATAGGGCAAAGCTAG-3'
   - Sense primer G: 5'-CAGGTGTACCCCTTTGATGACGAG-3'
   - Reverse primer H: 5'-CTCGTCATCAAAGGGGTACACCTG-3'

2.4.8. Transformation of plasmids in *E. coli*

2.4.8.1. Plasmid DNA preparation

Plasmid preparations were carried out using QIAGEN kits and following the manufacturer's instructions.

2.4.8.1a. Small scale preparation

For small-scale preparation or 'mini preps' the QIAprep Spin Miniprep kit (Qiagen) was used. Bacteria were picked from single colonies and grown in 2ml of LB broth containing ampicillin (or other antibiotics) overnight at 37°C in a bacterial shaker. 1.5ml of culture was transferred to an eppendorf tube and centrifuged at 5000 rpm for 5 min in bench top microfuge. The pellet was re-suspended in 250µl of buffer P1 (50mM Tris-HCl pH 8.0, 10mM EDTA, 100µg/ml RNAse A). 250µl of solution P2 (200mM NaOH, 1% SDS) was added and the tube was inverted several times to mix. To the same tube, 350µl of solution N3 (3.0M sodium acetate, pH 5.5) was added and immediately mixed by inverting the tubes 4-6 times. The mixture was centrifuged in a bench top microfuge for 10 min at 13,000rpm and the supernatant transferred to a QIAprep column. The column was centrifuged for 30 s at 13,000rpm in microfuge and the flow-through discarded. The columns were washed with 0.75ml of PE buffer (QIAprep Spin Miniprep kit, Qiagen) and the DNA was eluted with 50µl of ddH_2O.

2.4.8.1b. Large scale preparation

For large-scale plasmid preparations or 'maxi-preps' the QIAGEN Plasmid Maxi kit was used. Briefly, 250ml of LB containing 100µg/ml ampicillin was inoculated with 1ml of an overday culture of bacteria. The inoculated cultures were left to grow overnight at 37°C with vigorous shaking. Bacteria were harvested at 13,000g for 5 min at 4°C and the pellet was re-suspended in 10ml of re-suspension buffer P1 (50mM Tris-HCl pH 8.0, 10mM EDTA, 100µg/ml RNAse A). Ten ml of lysis buffer P2(200mM NaOH, 1% SDS) were added. After 5 min incubation at RT, 10 ml of ice cold neutralization buffer P3

(3mM potassium acetate pH 5.5) were added and the mixture was further incubated on ice for 20 min with occasional inversion. The cell debris was pelleted by centrifugation at 15000rpm in a Beckman JA-17 rotor for 30min. The supernatant was added to pre-equilibrate QIAGEN-tip 500 columns and allowed to enter the resin by gravity. The column was washed twice with 30 ml of wash buffer QC (1M NaCl, 50mM MOPS pH 7.0, 15% ethanol). The DNA was eluted with 15ml of elution buffer QF (1.25M NaCl, 50mM Tris-HCl pH 8.5, 15% Ethanol) and precipitated in 10.5ml of isopropanol at RT. Centrifugation was performed for 30 min at 15000rpm at 4°C in the same rotor previously used and the DNA pellet was washed with 70% ethanol. The pellet was air dried for 5 min and re-suspended in ddH_2O or Tris-EDTA buffer (10mM Tris-HCl, 1mM EDTA, pH 8).

2.4.8.2. Ethanol precipitation of DNA

Three volumes of 100% Ethanol and 1/10 volumes of 3M sodium acetate (pH 5.2) were added to the DNA samples. After mixing, the mixtures was kept on dry ice for 10 minutes. The sample was spun in a microfuge (top speed) at 4^0C for 30 mins. After decanting the supernatant, the pellet was dried by leaving the tube open for 15mins. To resuspend DNA, ddH_2O or TE was added followed by vortex.

2.5. RNA manipulation

2.5.1. Total RNA extraction

Total RNA was isolated from cells using the Trizol reagent (Invitrogen) as per the exact instructions of the manufacturer. Total RNAs were precipitated with isopropanol, washed with 70% ethanol, and dissolved in DEPC-treated H2O.

2.5.2. RNA quantification

The concentrations of RNA preparations were determined by measuring the optical density of the samples at 260nm (OD260) using NanoDrop (ThermoScientific). RNA concentration was calculated using the relationship: 1 OD unit at 260nm = 40 μg/ml RNA.

2.5.3. Reverse transcriptase PCR

2.5.3.1. First strand synthesis

Total RNA was quantified and 0.2–0.5 μg of RNA from each sample was used for the reverse transcription (RT) reaction. RNA was mixed with 5 μM random hexamers and the mix was heated to 70°C for 10 min after which 1 × first strand buffer, 10 mM dithiothreitol (DTT), 0.5 mM dNTPs and 0.75 U RNAaseOUT were added. After incubation at 20–22°C for 10 min and at 42°C for 2 min, 20 units of SuperScript® III Reverse Transcriptase (Invitrogen) were added and the reaction was incubated at 42°C for 50 min followed by 70°C incubation for 15 min.

2.5.3.2. Real-time quantitative PCR (RT-qPCR)

Real time PCR was performed with Lightcycler 480 (Roche) with LightCycler FastStart DNA Master^{PLUS} SYBR Green 1 (Roche). Gene-specific primer pairs were

designed and their specificity checked using the Universal ProbeLibrary Assay Design Center (Roche Applied Science at https://www.roche-applied-science.com/sis/rtpcr/upl/index.jsp?id=UP030000). Primers spanning intron/exon boundary were selected whenever possible to avoid amplification of contaminating genomic DNA. *Tbp* the TATA binding protein gene, was used as an internal control to normalize input RNA. All measurements were performed in triplicate. The transcript levels were quantified using the standard curve method. See Table 2.2 for sequences of primers used in RT-qPCR.

TABLE 2.2. Primer sequences

Gene	species	Forward primer	Reverse primer	RT-qPCR assay	Annealing temperature
Gsc	X.*laevis*	GATGCCGCCAGTGCCTC	TGCAGCTCAGTTCGTGACAAA	**SYBR**	**55^{0}C**
Eomes	X.*laevis*	ACCGGCACCAGACTGAGAT	TGAGGAAAGGGAACATCCTC	**SYBR**	**55^{0}C**
Fgf8b	X.*laevis*	GCAGAGCCTGGTGACCGA	TCTTGTTCGCCAGGATTTGC	**SYBR**	**55^{0}C**
Bmp4	X.*laevis*	AGCCCACTAAGGATGTGGTG	GCTGCTGAGGTTGAACACAA	**SYBR**	**55^{0}C**
Vent2	X.*laevis*	CCTCTGTTGAATGGCTTGCTT	GAGACTTGGGCACTGTCTG	**SYBR**	**55^{0}C**
Mixer	X.*laevis*	CACCAGCCCAGCACTTAACC	CAATGTCACATCAACTGAAG	**SYBR**	**55^{0}C**
Xbra	X.*laevis*	TTCTGAAGGTGAGCATGTCG	GTTTGACTTTGCTAAAAGAGACAGG	**SYBR**	**55^{0}C**
Odc	X.*laevis*	GCCATTGTGAAGACCTCTCTCCATTC	TTCGGGTGATTCCTTGCCAC	**SYBR**	**55^{0}C**
Pou5f1	M. *musculus*	GGCGTTCTCTTTGGAAAGGTGTTC	CTCGAACCACATCCTTCTCT	**SYBR**	**58^{0}C**
Nanog	M. *musculus*	ATGAAGTGCAAGCGGTGGCAGAAA	CCTGGTGGAGTCACAGAGTAGTTC	**SYBR**	**58^{0}C**
Klf4	M. *musculus*	CGGGAAGGGAGAAGACACT	GAGTTCCTCACGCCAACG	**SYBR**	**58^{0}C**
Sox2	M. *musculus*	GGCGGCAACCAGAAGAACAG	GCTTGGCCTCGTCGATGAAC	**SYBR**	**58^{0}C**
Rex1	M. *musculus*	ACGAGTGGCAGTTTCTTCTTGGGA	TATGACTCACTTCCAGGGGGCACT	**SYBR**	**58^{0}C**
Fgf4	M. *musculus*	CCGGTTCTTCGTGGCTATGA	CTTACTGAGGGCCATGAACATACC	**SYBR**	**58^{0}C**
E-Cad	M. *musculus*	AGACTTTGGTGTGGGTCAGG	CATGCTCAGCGTCTTCTCTG	**SYBR**	**58^{0}C**
Nodal	M. *musculus*	GGCGTACATGTTGAGCCTCT	GCCTGGTGGAAAATGTCAAT	**SYBR**	**60^{0}C**
Bmp4	M. *musculus*	CAACCCAATTATGGGCTGGC	CCACAATCCAATCATTCCAGC	**SYBR**	**58^{0}C**
TBra	M. *musculus*	GTGACTGCCTACCAGAATGA	ATTGTCCGCATAGGTTGGAG	**SYBR**	**60^{0}C**
Wnt3	M. *musculus*	CGCTCAGCTATGAACAAGCA	GGTGTTTCTCCACCACCATC	**SYBR**	**60^{0}C**
Tbx4	M. *musculus*	ACACCTTCCCAACTCAGAGG	CTTGCAGGGCAAGTCCAG	**SYBR**	**58^{0}C**
Sprouty2	M. *musculus*	GAGAGGGGTTGGTGCAAAG	CTCCATCAGGTCTTGGCAGT	**UPL (3)**	**58^{0}C**
Gata4	M. *musculus*	GCCTGCGGCCTCTACATGAA	CAGGACCTGCTGGCGTCTTA	**SYBR**	**60^{0}C**
Gata6	M. *musculus*	GGTCTCTACAGCAAGATGAATGG	TGGCACAGGACAGTCCAAG	**UPL(40)**	**58^{0}C**
Mixl1	M. *musculus*	AGTTGCTGGAGCTCGTCTTC	AGGGCAATGGAGGAAAACTC	**SYBR**	**60^{0}C**
Foxa2	M. *musculus*	CATCCGACTGGAGCAGCTA	GCGCCCACATAGGATGAC	**SYBR**	**58^{0}C**
Cdx2	M. *musculus*	GGAAGCCAAGTGAAAACCAG	CTTGGCTCTGCGGTTCTG	**SYBR**	**58^{0}C**
Tbp	M. *musculus*	GGGGAGCTGTGATGTGAAGT	CCAGGAAATAATTCTGGCTCA	**SYBR**	**58^{0}C**
Sdf2	M. *musculus*	TTCTCACGACGTGCGCTAT	TCATCCACAGAGGTCACACC	**SYBR**	**58^{0}C**
Msx2	M. *musculus*	AGGAGCCCGGCAGATACT	GTTTCCTCAGGGTGCAGGT	**UPL (70)**	**58^{0}C**
Fgfr1	M. *musculus*	CGAATTGGAGGCTACAAGGT	GAAGGCACCACAGAATCCAT	**UPL(1)**	**58^{0}C**
Otx2	M. *musculus*	GACTGCAGGGCAGAGACG	GGTAGATTTGGAGTGACGGAAC	**SYBR**	**58^{0}C**
Dkk1	M. *musculus*	CCGGGAACTACTGCAAAAAT	GGTTTTCAATGATGCTTTCCTC	**UPL (76)**	**58^{0}C**
Fgfr2	M. *musculus*	CCTGCGGAGACAGGTAACA	CGGGGTGTTGGAGTTCAT	**SYBR**	**58^{0}C**
Uhrf2	M. *musculus*	AAGGGGCAGTCAAAGAAGC	GGAGTCACCTGGACACTCATC	**SYBR**	**58^{0}C**

2.6. Protein analysis

2.6.1. Lysis and preparation of total protein extracts

Cells were washed twice with ice cold 1xPBS (Phosphate Buffered Saline (without $CaCl_2$ or $MgCl_2$), scraped into 1ml of ice cold 1x PBS and subsequently transferred to an eppendorf tube. The cells were centrifuged at 800 g at 4°C for 5 min and the pellets were immediately frozen at -70 °C until lysis was performed. Frozen pellets were lysed in 2 volumes of NP40 lysis buffer (1% (v/v) NonidetP-40, 100mM $NaCl_2$, 20mM Tris-HCl (pH 7.4), 10mM NaF, 1mM sodium orthovanadate, 30mM β-glycerophosphate and protease inhibitors ('Complete'protease inhibitor mixture, as instructed by the manufacturer (Roche). Cellular debris was spun down at 13,000 rpm (microfuge) for 10 min at 4°C. The lysate was transferred to a fresh tube to assay for protein concentration.

2.6.2. Determination of protein concentration

Protein concentrations were determined using the Bio-Rad D_c protein assay (Bio-Rad Laboratories), a protein assay based on the Lowry method (Lowry et al.,1951). Working reagent A was made by adding 20µl of reagent S to each one ml of reagent A as instructed by the manufacturer. A standard curve was established by assaying 5 dilutions of a protein standard (e.g. BSA) within the range of 0.2mg/ml to 1.5mg/ml protein. 100µl of reagent A' was added to 2µl of each dilution and subsequently mixed with 800µl of reagent B. Absorbance was read at 750nm after 15 min and a standard curve was established. Absorbance at 750nm was measured for each of the experimental samples and their concentrations were determined by multiplying the absorbance of the sample by the standard curve's regression coefficient.

2.6.3. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Gels for resolution of proteins were made by using a 30% (w/v) acrylamide/bis stock solution (containing a ratio of 29 acrylamide: 1 NN'-methylenebisacrylamide; Bio-Rad). As a polymerisation catalyst, 25% (w/v) ammonium persulphate (APS) and TEMED were used. For all percentages of resolving gels, a 4.5% (w/v) stacking gel was used.

SDS-PAGE gels were topped with 2cm of stacking gel (see Table 2.3). For separation of the proteins, 25µg of each cell lysate was mixed with 1 volume of 2x SDS loading buffer (4% (w/v) SDS, 62.5mM Tris-HCl (pH 6.8), 1% (v/v) glycerol, 0.01% (w/v) bromophenol blue, 10% (v/v) ß-mercaptoethanol) and heated at 90°C for 5 min. Samples were loaded into the stacking gel and fractionated by electrophoresis on SDS-PAGE gels (7%, 10% or 12.5%) at a constant voltage of 100V during the day in running buffer (25mM Tris-base, 190mM Glycine, 0.1% (w/v) SDS). Proteins were separated alongside broad-range pre-stained Rainbow SDS-PAGE standards (GE Healthcare)

TABLE 2.3 Proportions of the constituents of SDS-PAGE resolving and stacking gels.

Gel type	Resolving			Stacking
Percentage	5.00%	10.00%	12.50%	4.50%
30% Acrylamide (ml)	5	10	12.5	1.5
1M Tris (ml)	11.2 (pH 8.8)	11.2 (pH 8.8)	11.2 (pH 8.8)	1.25 (pH 6.8)
10% SDS (ml)	0.3	0.3	0.3	0.1
25% APS (µl)	50	50	50	25
TEMED (µl)	40	40	40	20
DH_2O (ml)	13.5	8.5	6.5	7.3

2.6.4. Western blotting of SDS-PAGE

Following separation via SDS-PAGE, proteins were transferred to a Protran nitrocelluose membrane (Whatman) by electrophoretic transfer in a wet tank blotting system (Bio-Rad Laborartories Trans-Blot cell). The transfer was carried out in transfer buffer (25mM Tris, 190mM glycine and 20% (v/v) methanol) for 1 h at a constant voltage of 100V at 4°C. The nitrocellulose membrane was blocked at RT by 1 h incubation in 5% (w/v) skimmed milk powder (Marvel) dissolved in Tris Buffered Saline/Tween (TBST) (20mM Tris-base, 136mM NaCl, adjusted to pH 7.6 supplemented with 0.05% Tween). The membrane was incubated for 2 h at RT (or overnight at 4°C) with the primary antibody (1:1000 dilution) of choice (see Table 2.1) diluted in 5% (w/v) Marvel/TBS-T at dilutions recommended by the supplier. The membrane was washed five times (10 min each at RT) in TBS-T prior to incubation with

horseradish peroxidase (HRP) conjugated secondary anti-mouse, anti-rabbit or anti-goat conjugates as appropriate (GE Healthcare). The membrane was exposed to an auto-radiographic film for times varying from 10 s to 1h and visualized using the ECL-plus detection system (GE Healthcare).

2.7. Microarray analysis

RNA was prepared from different ES cell lines as discussed above. The RNA was sent to collaborators, Minoru Ko and Alexei Sharov at the NIH in Baltimore for further analysis. Biological and technical replicates for each sample were hybridised to NIA Mouse 44K Microarray v2.1 according to the published protocols (Carter et al., 2005, Sharov et al., 2008)). Briefly, "2.5 μg of total RNA was labeled with Cy3-CTP using a Low RNA Input Fluorescent Linear Amplification Kit (Agilent,). A reference target (Cy5-CTP-labeled) was prepared from the Universal Mouse Reference (UMR) RNA (Stratagene). Labeled samples were purified using an RNeasy Mini Kit (Qiagen, USA) according to the Agilent's protocol, quantified by a NanoDrop scanning spectrophotometer (ThermoScientific), and "hybridized to the NIA Mouse 44 K Microarray v2.1 (whole genome 60-mer oligo; manufactured by Agilent Technologies, #012799) according to the Agilent protocol (G4140-90030; Agilent 60-mer oligo microarray processing protocol – SSC Wash, v1.0)" (from Sharov et al., 2008). All hybridizations were carried out in the two color protocol by combining one Cy3-CTP-labeled experimental target and Cy5-CTP-labeled reference target. Microarrays were scanned on an Agilent DNA Microarray Scanner, using standard settings. Further discussions of the protocols are detailed in Chapter 3.

2.7.1. Statistical analysis of microarrays

Statistical analysis was done using the NIA Array Analysis software (Sharov et al., 2005). A. Sharov using standard statistical conditions (FDR<0.05, 2-fold expression levels change) to unveil genes with changes in expression levels between the samples and controls.

Gene ontology (GO) analysis was done using the NIA Mouse Gene Index software, which evaluates statistical significance using the hypergeometric probability distribution with parameters: FDR = 0.05. (see Sharov etal., 2008 & http://lgsun.grc.nia.nih.gov/geneindex/mm8/).

CHAPTER 3

Oct4 activation function is sufficient to block lineage commitment and maintain ES cell self-renewal

3.1. Introduction

In vertebrate development, lineage commitment occurs throughout gastrulation, a complex and co-coordinated series of cellular movements that progress in an anterior to posterior direction.

Although the precise choreography of morphogenic movements and rearrangements varies between species, germ layer specification occurs throughout gastrulation in all species and requires a pool of multipotent uncommitted progenitor cells to complete this process. The need to maintain this multipotent uncommitted population is common to all species in which development and differentiation occur over time. In mammals, the molecular mechanism(s) that regulate this process have been exploited to maintain progenitor populations over longer window of time in order to generate the additional extra-embryonic lineages. In addition to its role in early mammalian development, the network maintaining progenitor cells also supports ES cell self-renewal and pluripotency.

One of the core components of the network responsible for regulating both ES cell pluripotency and progenitor maintenance in embryos is the Class V POU domain transcription factor Oct4.

In embryonic development, the conserved role of Oct4 related proteins is to regulate differentiation in a number of different progenitor populations. In mouse, Oct4 is expressed from the beginning of development in the unfertilised egg and the early blastocyst before becoming confined to the ICM (Pesce and Scholer, 2001). Oct4 mRNA persists in the pluripotent epiblast of the pre- and post-implementation embryo, and becomes progressively restricted to a posterior region in which cells are still undergoing mesoderm/endoderm induction and ingressing through the streak and the node. It is then downregulated, except in the primordial germ cells during their migration, and the formation of the genital ridges in both sexes (Palmieri et al., 1994). Oct4 null embryos die at peri-implantation stages due to the absence of Inner Cell Mass (ICM) and an expansion of trophectoderm (Nichols et al., 1998). Oct4 is also essential for ES cell self-renewal as its knockdown results in differentiation towards both primitive endoderm and trophoblast (Hay et al., 2004).

The relationship between the activity of Oct4 in ES cells and its putative role in gastrulation is apparent, as PouV proteins from vertebrate species lacking extra-embryonic development are able to maintain murine ES cells in the absence of the endogenous Oct4 (Morrison and Brickman, 2006). In fact, one of the Xenopus PouV proteins, Xlpou91, can do so as efficiently as the mouse protein. In Xenopus, knockdown of the PouV proteins results in precocious gastrulation stage differentiation, which can be rescued by expression of Oct4. Interestingly, while Oct4 activity is conserved in some vertebrates, it is not universally well conserved. For example, Zebrafish DrPou2 has little ability to rescue either PouV depleted Xenopus embryos or substitute for Oct4 in murine ES cells (Morrison and Brickman, 2006).

Understanding the mechanism by which PouV proteins block differentiation in the majority of vertebrates is vital. In Xenopus, PouV proteins appear able to maintain multipotent populations by suppressing differentiation during germ layer specification at least partly through inhibition of the Nodal related TGF-ß signaling (Morrison and Brickman, 2006). These findings led to the suggestion that Oct4 might regulate differentiation by directly suppressing Nodal and Wnt responsive targets.
In fact, it was proposed that a simple mode by which Oct4 could suppress differentiation in general would be the direct repression of lineage specific promoters. In mouse and human ES cells the inspection of genome wide Oct4 target screens suggested that this form of "lineage repression," was an important component of Oct4 activity (Boyer et al., 2005); (Loh et al., 2006). These studies proposed that Oct4 and its partners both activate genes encoding components of key signalling pathways, and repress differentiation specific genes in undifferentiated cells to maintain their self-renewal (Boyer et al., 2005); (Loh et al., 2006). This model was also based on earlier findings proposing that Oct4 is able to act as a repressor or as an activator of target genes (Pesce and Scholer, 2001).

In this chapter, we test whether direct lineage repression is an essential component of the ability of Oct4 to regulate differentiation in embryonic populations and maintain their self-renewal.

Here we generate Oct4 and Xlpou91 fusion proteins that function exclusively as activators or repressors and ask whether these proteins can block differentiation in embryonic cell populations and maintain their self-renewal.

We found that fusion proteins behaving as activators can block differentiation of both murine ES cells and embryonic progenitor cells in Xenopus. However, repressor proteins appear to induce differentiation in both models. Interestingly, we can derive cell lines in which Oct4 activator form expression replaces that of Oct4. These cell lines express high levels of pluripotency genes and no longer differentiate in the presence of high levels of Oct4 expression or in LIF withdrawal conditions. Moreover, analysis of these cell lines by microarray suggests that predicted Oct4 repressed targets remain repressed, and that upregulated and downregulated gene-sets in these cell lines correlate very closely with the global sets of genes up and down regulated by Oct4. Finally, we were also able to titrate down the levels of activated Oct4λVP2 in these cell lines, causing them to regain their ability to differentiate in the absence of LIF. Based on a preliminary analysis of our complete set of lines, it would appear that the major player causing the change of this differentiaton phenotype is levels of Nanog.

3.2. Results

3.2.1. Converting Oct4 and Xlpou91 to either activators or repressors

To test the importance of transcriptional activation and repression by PouV proteins in maintaining uncommitted cells during gastrulation and ES cell self-renewal, we constructed a series of global activator and repressor fusion proteins. We employed reiterated modular domain that have the capacity to generate an enhancer or a silencer bound to a single DNA site (Figure 3.1) These reiterated modules were originally tested in the context of Gal4 fusion proteins (Carey et al., 1990).

These modular units employ either the minimal activation domain of herpes simplex activator VP16 or the defined repression domain of the drosophila engrailed protein. Each domain is fused to the flexible hinge region of lambda C1 to create a protein sequence with multiple regulatory domains and the flexibility normally imparted from multiple DNA sites (Croston et al., 1992; Emami and Carey, 1992; Ohashi et al., 1994). The same strategy was used for both Oct4 and Xlpou91 (Figure 3.1). This approach has been effective for the generation of similar functional fusions with the homeodomain proteins Hex and Hesx1 (Brickman et al., 2000).

VP16 (65-KD) is a herpes simplex virus type 1 (HSV-1) protein, it induces the transcription of the immediate early (IE) genes during the lytic infection by the virus (Campbell et al., 1984). Its transcriptional activation domain is a highly acidic stretch in the carboxyl-terminal 78 amino acid of the protein (Triezenberg et al., 1988a; Triezenberg et al., 1988b), and was reported to be a very potent activator of transcription in mammalian cells when fused to the yeast activator GAL4 DNA binding domain (Sadowski et al., 1988). This VP16 activation domain is composed of two transactivation regions; region 1 (amino acids 411 to 456) and region 2 (amino acids 456 to 490) (Triezenberg et al., 1988a; Triezenberg et al., 1988b). While both regions have transactivation ability when fused to Gal4, region 1 was shown to be the most critical for the transactivation by full length VP16 (Sadowski et al., 1988) (Carey et al., 1990) (Croston et al., 1992; Emami and Carey, 1992; Ohashi et al., 1994). In fact, multimerization of VP16 region 1 was an effective approach to generating synergistic increases in activator potency, and the inclusion of a flexible

linker sequence that allows lambda C1 to bind to non-adjacent sites enhances this synergy without directly contributing to the activation potential of these molecules (Croston et al., 1992; Emami and Carey, 1992; Ohashi et al., 1994) (Figure 3.1a2, and b2).

Based on the ability to use domain synergy for activation, we took similar approach to generate the repressor constructs. The repression domains employed were derived from the Drosophila homoedomain protein Engrailed (EnR). EnR has the capacity to repress both basal and activator enhanced transcription trough binding to its specific targets (Han and Manley, 1993a; Han et al., 1991; Jaynes and O'Farrell, 1991). Repression of these targets is dependent on an amino terminal repression domain that can function when fused to a heterologous DNA binding domain (Jaynes and O'Farrell, 1991). This repression domain appears to contain two essential motifs. The first one is an alanine rich domain from residues 228 to 282. It is essential for EnR repression in vitro and can function as portable repressor domain when fused to a heterologous DNA binding domain (Han and Manley, 1993a; Tolkunova et al., 1998). The second domain eh1 (amino acids 172 to 186) represents a conserved motif present in a number of homeodomain proteins (Smith and Jaynes, 1996), and its function is more relevant to EnR mediated repression in vivo through recruitment of the Groucho co-repressor (Han and Manley, 1993b; Tolkunova et al., 1998). Our EnR sequence employed the combination of these two motifs (amino acids 168-281).

Oct4 and Xlpou91 activation and repression domain fusions are depicted in Figure 3.1a1 and b1.

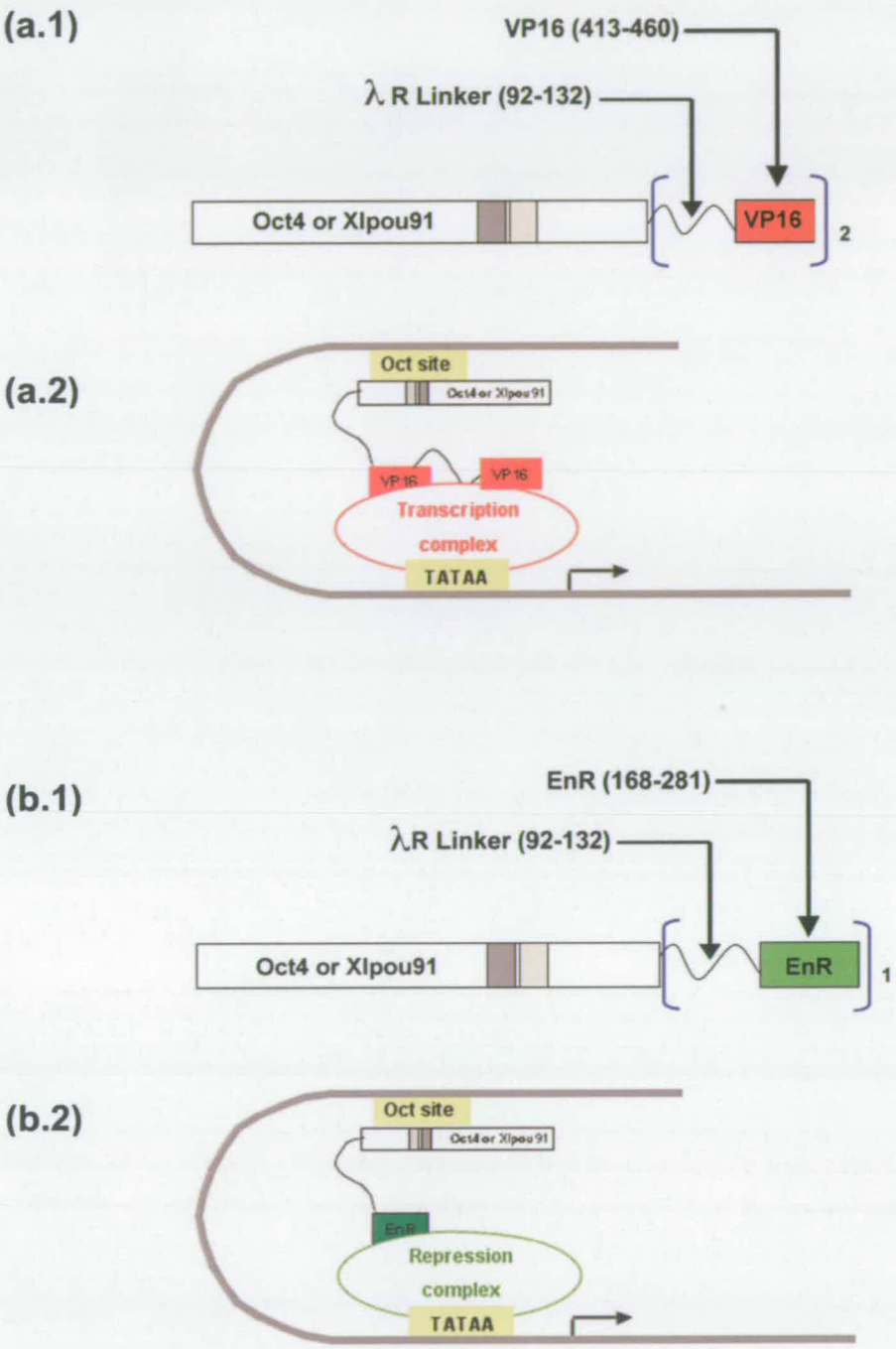


Figure 3.1-Schematic representation of the activator and the repressor forms of Oct4 and Xlpou91

(a.1) Diagram illustrating the construction of the activator form of Oct4 and Xlpou91proteins.
(a.2) Sketch illustrating the principle behind the design of Oct4/Xlpou91λVP2 fusions.
(b.1) Diagram illustrating the construction of the repressor form of Oct4 and Xlpou91 proteins.
(b.2) Sketch illustrating the principle behind the design of Oct4/Xlpou91λEnR fusions

To confirm that these fusion proteins behaved as only activators or repressors in ES cells, they were transiently cotransfected alongside luciferase reporter genes into the Tc suppressible ZHBTc4 ES cells in the presence of tetracycline (such that there was no endogenous Oct4 expression) (Niwa et al., 2002). The reporter genes assayed all employed a minimal thymidine kinase (*tk*) promoter driving luciferase downstream of Oct4 dependent sequences either from the *Fgf4* enhancer or 6 reiterated copies of the octamer-binding motif (Niwa et al., 2002). The basal level transcription from the *tk* promoters can either be activated or repressed (Figure 3.2a).
Moreover, the specificity of these fusion proteins was tested using a luciferase reporter lacking any octamer-binding sites (Figure 3.3a).

Figure 3.2b shows that both Oct4λVP2 and Xlpou91λVP2 function as potent activators of gene transcription from both the *Fgf4* enhancer and the octamer-binding reporters. On the other hand, Oct4λEnR and Xlpou91λEnR repressed reporter gene transcription (Figure 3.2c).
The specificity of the transcriptional activity of these fusion proteins was confirmed as no induction of transcription was observed from the tk promoter luciferase reporter lacking any PouV protein responsive elements (Figure 3.3b and c).

Oct4 has been shown to repress transcription of β-Subunit of Human Chorionic Gonadotropin (*hCG-β*) promoter in human choriocarcinoma Jar cells (Liu et al., 1997). To confirm that Oct4λVP2 was unable to repress transcription in these cell lines, Oct4λVP2 was co-transfected alongside an *hCG-β* promoter into Jar cells and levels of relative transcription determined. Figure 3.4 demonstrates that Oct4λVP2 was unable to repress *hCG-β* transcription, and therefore is unable to recapitulate the inhibition of *hCG-β*-promoter by wild type Oct4.

This data indicates that Oct4λVP2 appears to be an activator in all tested ES cell contexts, and to have no residual repression activity.

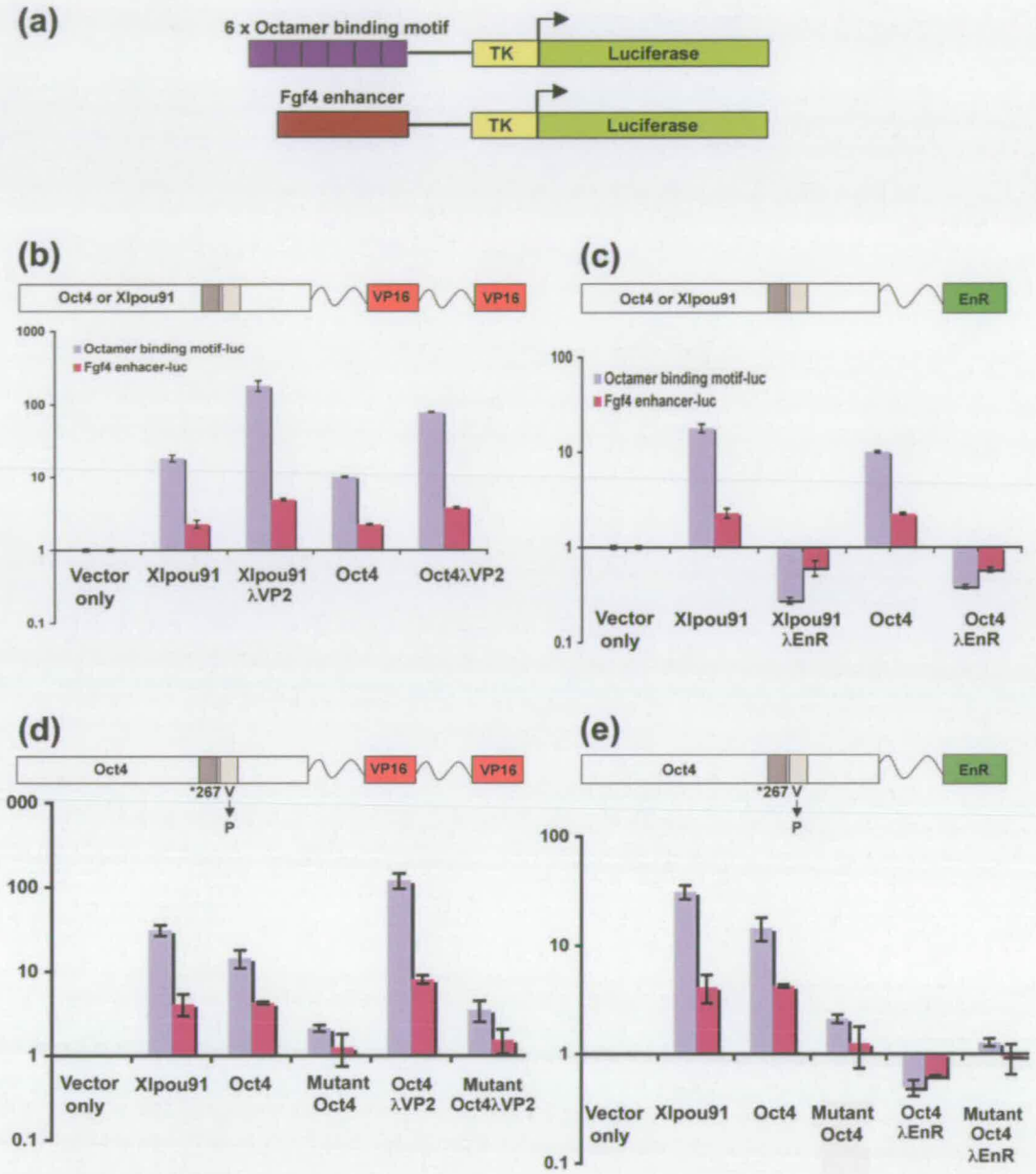


Figure 3.2-Transcriptional activity of the activator and the repressor form of Oct4 and Xlpou91

(a) Schematic representation of the luciferase reporters used to test transcriptional activity of the activator and repressor forms of Oct4 and Xlpou91: The Fgf4 enhancer and the octamer binding motif luciferase reporters.
(b) Oct4 and Xlpou91 activator forms are able to activate transcription from two different luciferase reporters
(c) Oct4 and Xlpou91 repressor forms are able to repress transcription from two different luciferase reporters
(d) DNA binding mutant activator form of Oct4 still has some residual activity
(e) DNA binding mutant repressor form of Oct4 retains some residual activity.
The indicated cDNAs were cotransfected with the Fgf4 enhancer or the octamer binding motif luciferase reporters. Fold induction represents the increase in transcription compared with the vector only control. Data represent the mean value of two independent experiments.

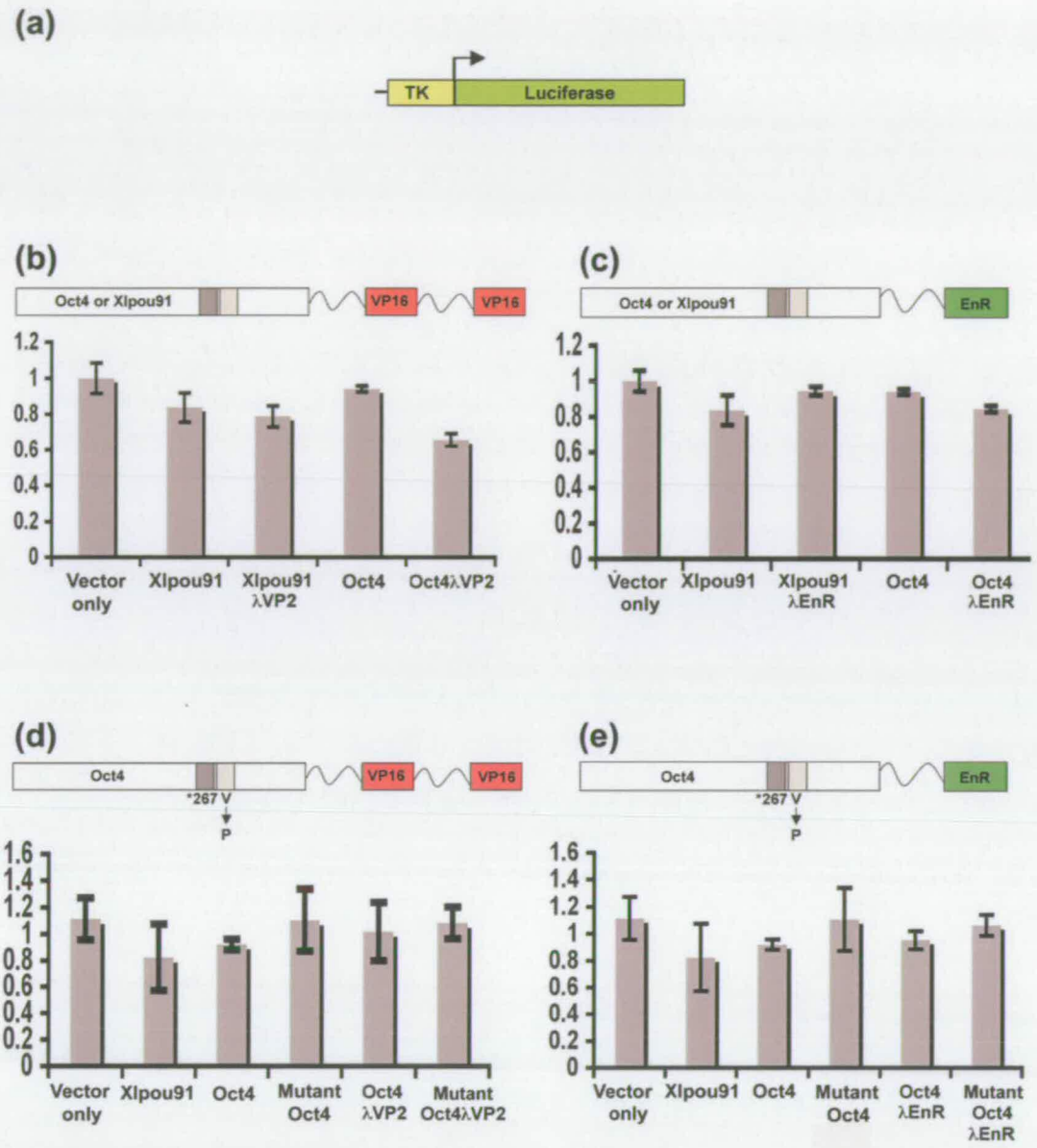


Figure 3.3-Transcriptional activity of the activator and the repressor form of Oct4 and Xlpou91 is specific

(a) Schematic representation of the luciferase reporter used to test the specificity of Oct4 and Xlpou91 activator and repressor forms transcriptional activity: The thymidine kinase (TK) promoter luciferase reporters.
(b) Oct4 and Xlpou91 activator forms show background transcriptional activity similar to the vector only activity.
(c) Oct4 and Xlpou91 repressor forms show background transcriptional activity similar to the vector only activity.
(d) DNA binding mutant activator form of Oct4 show background transcriptional activity similar to the vector only activity.
(e) DNA binding mutant repressor form of Oct4 show background transcriptional activity similar to the vector only activity.
The indicated cDNAs were cotransfected with the pTK promoter luciferase reporters. Fold induction represents the increase in transcription compared with the vector only control. Data represent the mean value of two independent experiments.

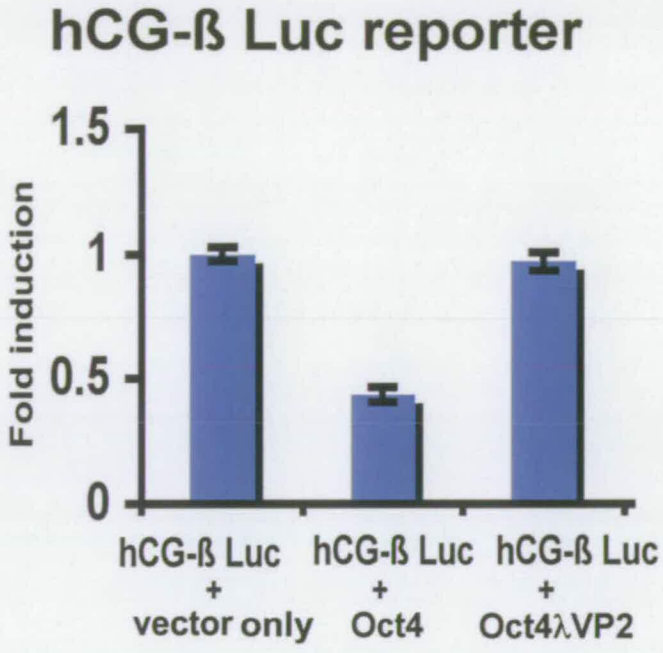


Figure 3.4-Oct4 activator form is unable to repress transcription from the β-Subunit of Human Chorionic Gonadotropin (hCG-β) promoter

The hCG-β-Luciferase construct (2ug) was cotransfected into Jar cells with 2ug of either pCAGIP (vector only as a negative control), CAG-Oct4, CAG -Oct4λVP2.

Oct4λVP2 is unable to recapitulate the inhibition of hCG-β-Luciferase by Wild type Oct4.

Fold induction represents the increase or decrease in transcription compared with the vector only control.

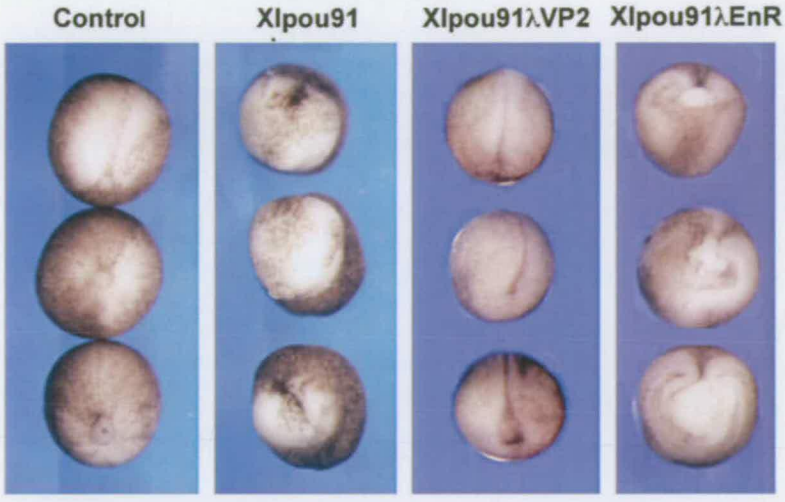
3.2.2. Activation by Oct4 is sufficient to repress lineage specific differentiation in Xenopus embryos

As it has been suggested that Oct4 and its homologues directly suppress gastrulation stage targets of the Nodal and Wnt pathway such as *Goosecoid* as a mean to regulate differentiation, we asked whether PouV fusion proteins (Oct4λVP2 or Xlpou91λVP2) acting as activators only could suppress differentiation in Xenopus embryos.

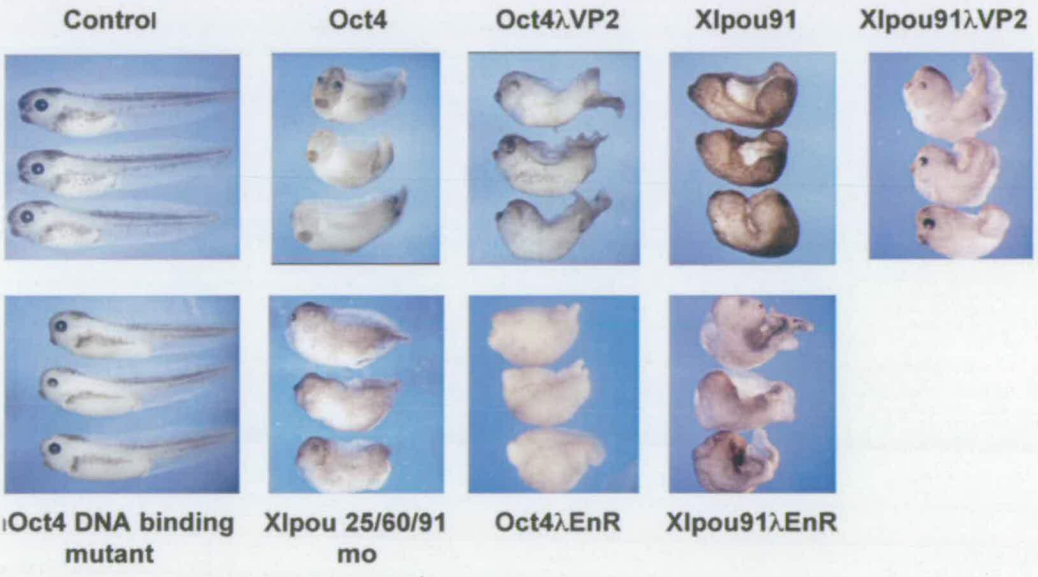
Figure 3.5a1 and a2 show the overexpression phenotype of the activator and repressor forms of Oct4 and Xlpou91 compared to the phenotype of embryos depleted of all three Xenopus PouV proteins. Injection of RNA encoding *Oct4* or *Xlpou91* produces complex axis truncations, however embryos still appear to exhibit an overt anterior-posterior axis polarity. In the case of *Xlpou91*, we also observe a failure in neural tube closure and exaggerated pigmentation (Figure 3.5a2). However, deletion of the three PouV proteins by Morpholino Oligos (MO) produces defects in both the head and the tail alongside an axis truncation. Interestingly, phenotypes obtained from injection of RNA encoding *Oct4λVP2* and *Xlpou91λVP2* appear to induce phenotypes that superficially resemble the overexpression phenotype of the wild type proteins Oct4 and Xlpou91 rather than the PouV protein MO knockdown phenotype. In fact, while all activators overexpression phenotypes show no reduction in the anterior character, overexpression of the EnR fusions exhibit profound anterior and posterior truncations and appears to better recapitulate the knockdown phenotype at least on a superficial phenotype level.

Unexpectedly, there is one aspect of wild type PouV overexpression phenotypes that appears linked to repression. In the *Xlpou91λEnR* injected embryos there appears some upregulation of pigmentation, while not as significant as that observed with *Xlpou91* the significance of this remains unclear. These crude phenotypes are suggestive of a positive role for PouV proteins in regulating gene expression, however to confirm these observations we needed to look at specific markers regulated by PouV proteins in gastrulation. Figure 3.5b1 and b2 show the expression of a number of gastrulation stage markers as determined by both in situ hybridisation and quantitative RT-PCR on RNA from stage 10 whole embryos.

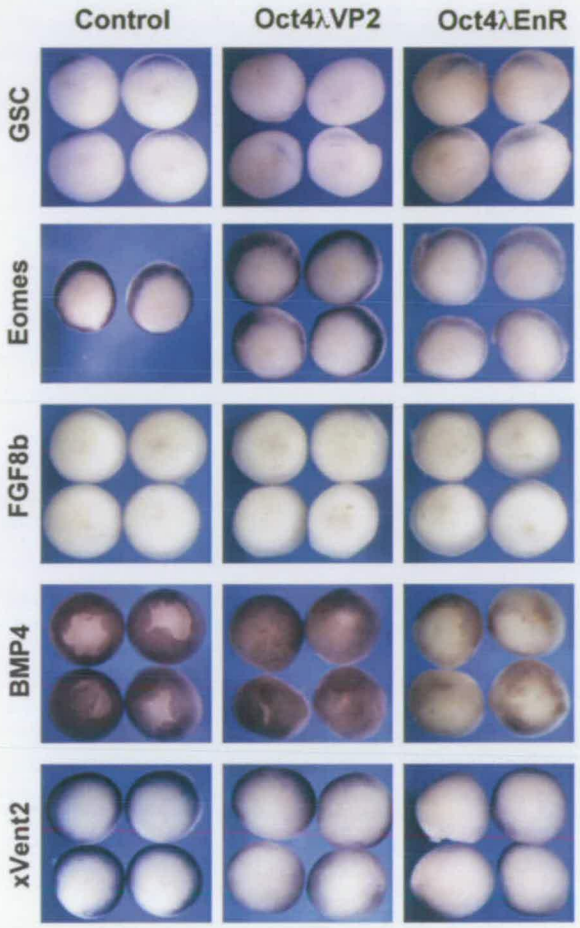
(a.1)


Control
Xlpou91
Xlpou91λVP2
Xlpou91λEnR

(a.2)


Control
Oct4
Oct4λVP2
Xlpou91
Xlpou91λVP2
Oct4 DNA binding mutant
Xlpou 25/60/91 mo
Oct4λEnR
Xlpou91λEnR

(b.1)


Control
Oct4λVP2
Oct4λEnR
GSC
Eomes
FGF8b
BMP4
xVent2

(b.2)

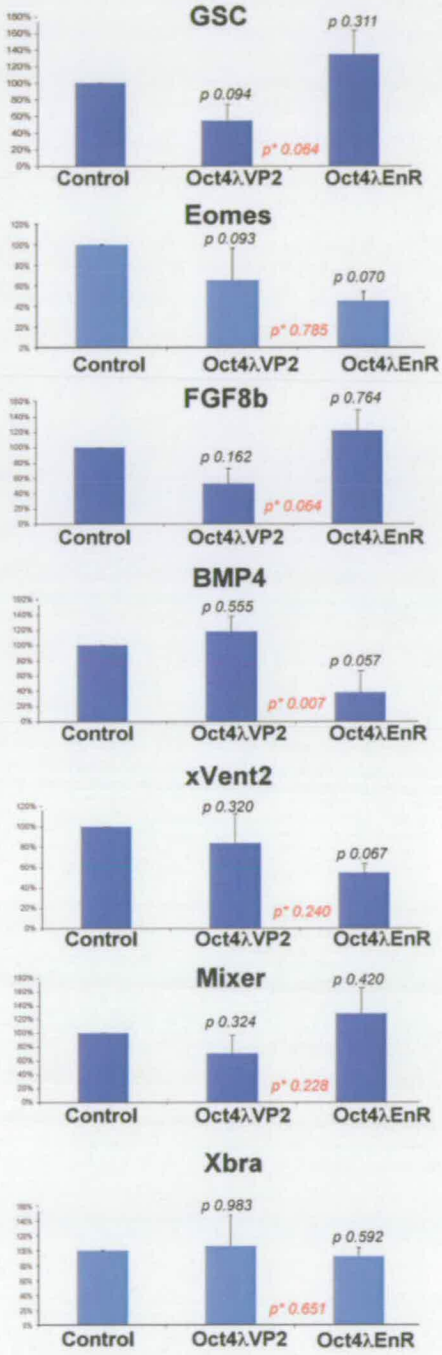

GSC
p 0.311
p 0.094
p* 0.064
Control
Oct4λVP2
Oct4λEnR
Eomes
p 0.093
p 0.070
p* 0.785
Control
Oct4λVP2
Oct4λEnR
FGF8b
p 0.764
p 0.162
p* 0.064
Control
Oct4λVP2
Oct4λEnR
BMP4
p 0.555
p 0.057
p* 0.007
Control
Oct4λVP2
Oct4λEnR
xVent2
p 0.320
p 0.067
p* 0.240
Control
Oct4λVP2
Oct4λEnR
Mixer
p 0.420
p 0.324
p* 0.228
Control
Oct4λVP2
Oct4λEnR
Xbra
p 0.983
p 0.592
p* 0.651
Control
Oct4λVP2
Oct4λEnR

Figure 3.5-Activation by Oct4 is sufficient to repress lineage specific differentiation in *Xenopus* embryos

Overexpression phenotypes of activator and repressor forms of Oct4 and Xlpou91.

(a.1) Overexpression phenotype of activator and repressor forms of Oct4 and Xlpou91 in stage 12.5 *Xenopus* embryo. Embryos were injected at two-cell stage in both blastomeres with 250 pg of mRNA
(a.2) Ovexpression phenotype of activator and repressor forms of Oct4 and Xlpou91 in *Xenopus* embryo at stage 39. Embryos were injected (same as a.1).

(b.1) Effect of overexpressing Oct4 activator and repressor forms on lineage specific gene expression. *In situs* were performed on stage 10.25 embryos following injection of mRNA (same as a.1). Pictures are taken of the vegetal pole except BMP4, which is an animal view.

.

(b.2) qRT-PCR analysis of total RNA collected from stage 10.25 embryos following injections of mRNA (same as a.1). The relative change in gene expression is shown as a percentage of that from the control embryos. These values are the average of three independent experiments.
Unpaired two tailed equal variance student t-test was used to compare gene expression levels between the control embryos and embryos overexpressing either Oct4 activator or repressor form (*p* values in black).
It was also used to determine the statistical significance of the gene expression differences between embryos overexpressing Oct4 activator form, and embryos expressing its repressor form (*p* values in red).

These *Xenopus* experiments were done in collaboration with Dr Gillian Morrison, and Dr Allessandra Livigni.

Interestingly, organizer markers such as *Goosecoid*, and *Mixer* are suppressed by Oct4λVP2 and induced by Oct4λEnR indicating an indirect mechanism of lineage repression. However, marginal zone markers previously shown to be PouV dependent such as *Bmp4*, the Bmp target *Xvent2* or Eomes are suppressed by Oct4λEnR. Taken together, this data indicates that the induction of endoderm and organizer specification is suppressed by a network downstream of the PouV proteins rather than a direct repression by PouV proteins.

An interesting candidate member of this network that appears highly responsive to expression of Oct4λVP2, Oct4λEnR or morpholinos mediated depletion of PouV proteins is *Bmp4* (Morrison and Brickman, 2006). Bmp4 has been associated with maintaining non-committed cells in both ES cells and in Xenopus embryos as expression of its constitutively active receptor blocks the onset of germ layer specification (Constance Lane et al., 2004; Ying et al., 2003). As these observations suggest that the loss of *Bmp4* expression could explain a number of the phenotypes observed in the PouV depleted embryos, we investigated this possibility by testing the capacity of *Bmp4* expression to rescue the PouV depletion phenotype. While not tested extensively, injection of *Bmp4* RNA alongside morpholinos to the PouV proteins appears to rescue the expression of both *Xvent2* and *Gsc* (Figure 3.6).

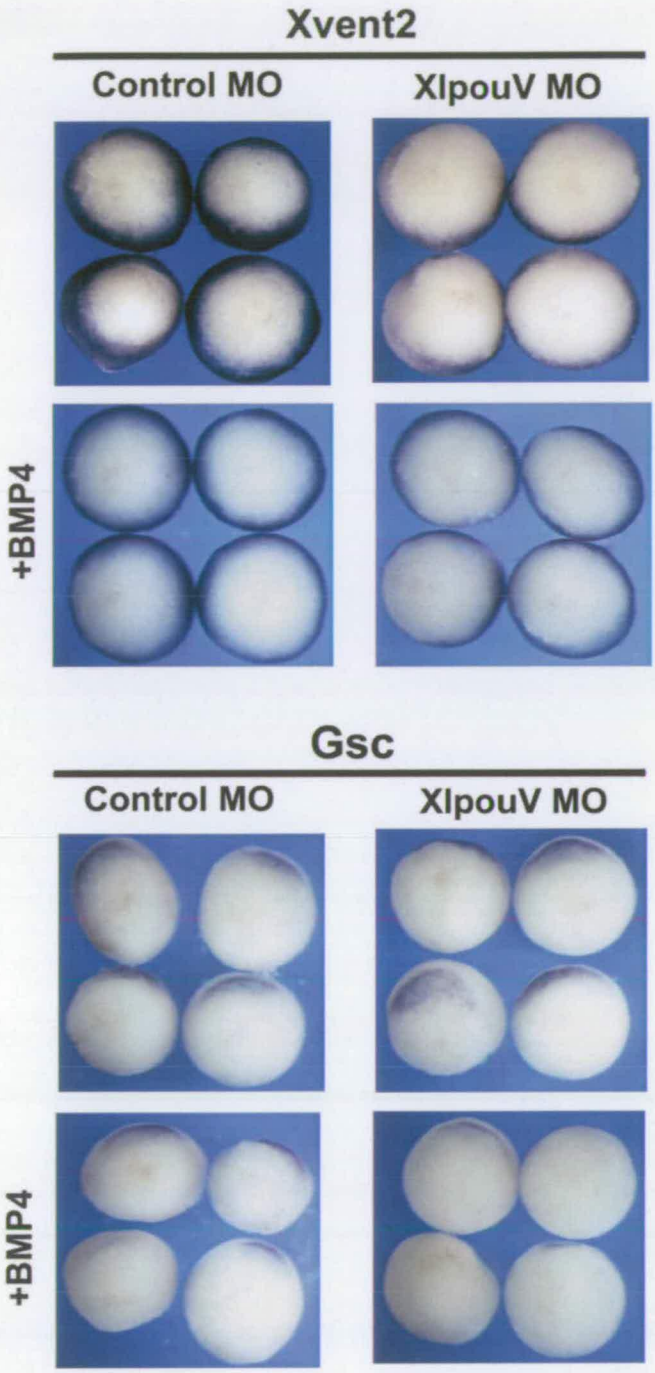


Figure 3.6-BMP4 mRNA rescues gene expression of Xl pouV target genes following XlpouV deplitions in Xenopus embryos

In situ hybridisation analysis of stage 10.25 embryos following injection of control morpholino, PouV Morpholino, or coinjection of BMP4 mRNA with each morpholino treatment.

This experiment was done in collaboration with Dr Allssandra Livigni, and Lucy Jones.

3.2.3. Activator function of Oct4 is sufficient to rescue ES cell self-renewal

As activation by PouV proteins was found to be sufficient to repress lineage specific differentiation in embryonic progenitors in Xenopus embryos, we wished to determine whether it was also sufficient to support ES cell self-renewal. We tested the ability of the activator and repressor forms of Oct4 and Xlpou91 proteins to rescue ES cell self-renewal using the complementation assay designed by Niwa and colleagues (Niwa et al., 2002) (Figure 3.7).

This assay uses the ZHBTc4 cell line whereby both alleles of endogenous Oct4 were disrupted by homologous recombination. The cells are maintained by a randomly integrated tetracycline-regulated Oct4 transgene that can be repressed by the addition of tetracycline (Tc) to the medium. The system is based on the fact that in the absence of Oct4 (presence of Tc) ZHBTc4 cells differentiate, and that this phenotype can be rescued by introduction of a second Oct4 transgene via random integration. Therefore, the relative ability of different PouV proteins to substitute for Oct4 can be evaluated based on their capacity to rescue ES cell growth when introduced into ZHBTc4 cells in the absence of Oct4 (presence of Tc).

The assay was performed by electroporating $2x10^7$ ZHBTc4 ES cells with 100μg of linearised vector DNAs. To control for transfection efficiency, each transfection is divided in half and only one half is treated with Tc. Two days after the transfection, puromycin selection was applied for 9 days, and the cells were then stained with alkaline phosphatase (AP). The relative ability of a PouV protein to rescue ES cell self-renewal can be quantified by calculating a rescue index, that is the number of rescued alkaline phosphatase (AP) positive ES cell colonies obtained in the presence of Tc normalized to the number obtained in its absence for a given transfection. The values were then all normalized to that obtained with Oct4.

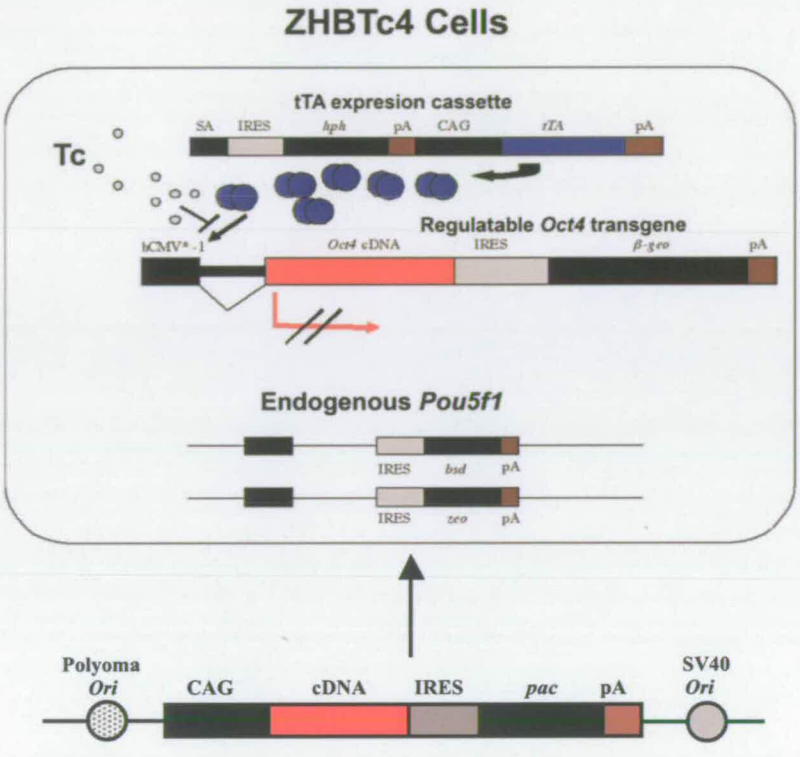


Figure 3.7- ZHBTc4 cell s complementation assay

ZHBTc4 cells have both alleles of endogenous *Pou5f1* disrupted by homologous recombination. The cells are maintained by a randomly integrated tetracycline regulatable Oct4 transgene that can be repressed by the addition of tetracycline (Tc) to the medium.

The system is based on the fact that in the absence of Oct4 (presence of Tc) ZHBTc4 cells differentiate to TE lineage. If Oct4 is supplied via introduction of a second transgene, this phenotype can be rescued and normal alkaline phosphatase positive ES cells can grow. Thus, the relative ability of different PouV proteins to substitute for Oct4 can be evaluated based on their capacity to rescue ES cell growth when introduced into ZHBTc4 cells in the absence of Oct4 (presence of Tc).

Figure 3.8 shows that the activator fusion proteins have the capacity to support undifferentiated ES cell growth in the absence of endogenous Oct4, whereas the repressor forms of the proteins do not. The rescue index obtained from transfection of activator fusions ranged from 20-70% of that obtained with the mouse Oct4, whereas PouVλEnR fusions produced no colonies in the absence of Oct4 and induced differentation in its presence (Fig 3.8a, and b3). As reported previously and demonstrated elsewhere in this thesis, Xlpou91 and Oct4 have similar activity in this assay.

The morphologies of representative colonies rescued by various control PouV proteins in the presence and absence of Tc regulatable Oct4 gene expression are shown in Figure 3.8 b1. Colonies rescued with Xlpou91 protein are AP positive, with undifferentiated ES cell morphology similar to those rescued with Oct4. These colonies generally exhibit undifferentiated AP growth in the centre with differentiated cells around the periphery. In the absence of Tc, the colonies generated by Oct4 and Xlpou91 appear more differentiated as a result of the overexpression of PouV protein (protein from both the Tc regulatable Oct4 transgene and the transfected PouV cDNA) (Figure 3.8b1).

Colonies rescued by either Oct4λVP2 or Xlpou91λVP2 exhibit a perfect undifferentiated ES cell morphology, and are positive for Alkaline phosphatase (AP) staining throughout (Figure 3.8b2). Moreover, the morphologies of colonies generated by these activator forms in the presence of Tc regulatable Oct4 transgene expression contrast greatly to those generated in the same conditions by wild type Oct4 and Xlpou91. This observation suggests that they are resistant to the Oct4 dose dependent differentiation response observed normally (Niwa et al., 2002).

While no colonies were obtained when the λEnR fusions were used to rescue Oct4 null cells, extremely low number of differentiated colonies was obtained when these fusions were expressed in the presence of the Tc regulatable *Oct4* transgene expression (Figure 3.8b3).

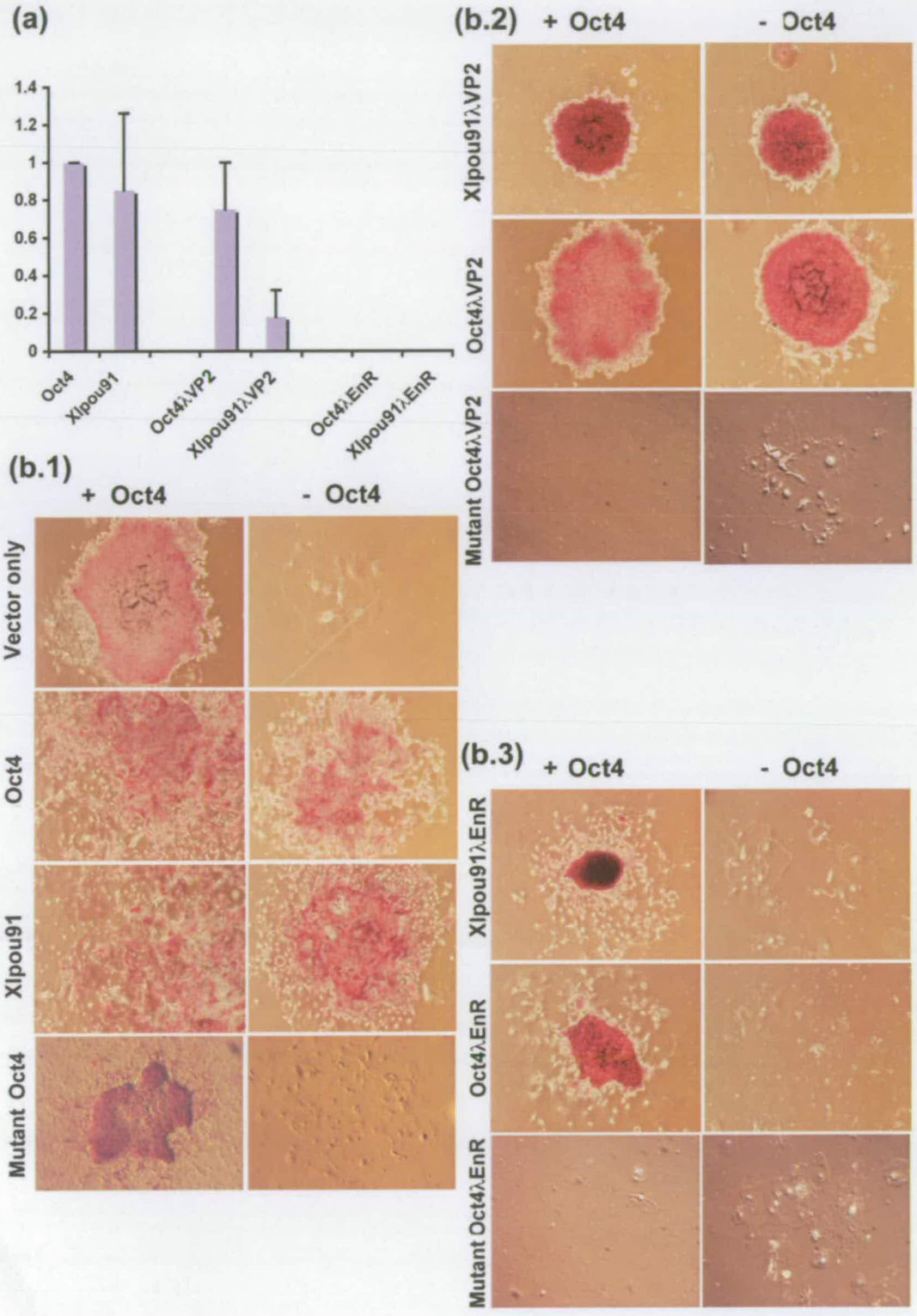

(a)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Oct4
Xlpou91
Oct4λVP2
Xlpou91λVP2
Oct4λEnR
Xlpou91λEnR
(b.1)
+ Oct4
- Oct4
Vector only
Oct4
Xlpou91
Mutant Oct4
(b.2)
+ Oct4
- Oct4
Xlpou91λVP2
Oct4λVP2
Mutant Oct4λVP2
(b.3)
+ Oct4
- Oct4
Xlpou91λEnR
Oct4λEnR
Mutant Oct4λEnR

Figure 3.8-Activator function of Oct4 is sufficient to rescue ES cell self-renewal

2x107 ZHBTc4 ES cells were transfected with 100ug of linearized vector DNAs by electroporation, after 9 days of selection with puromycin in the presence or absence of Tetracycline, the cells were stained with alkaline phosphates (AP) and the number of AP positive stem cell colonies were scored, and the rescue index calculated.

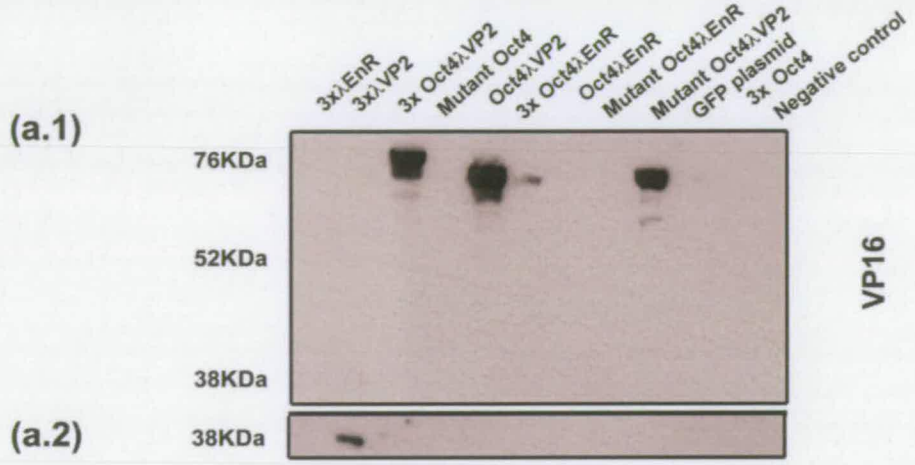
(a) Rescue index for PouV activator and repressor forms. The Rescue index is calculated by dividing the number of ES cell colonies in the absence of Oct4 (presence of Tc) by the number of AP positive ES cell colonies present in the presence of Oct4 (absence of Tc). Rescue index values were normalised to Oct4 value. Data represents the mean values obtained from three independent experiments.
(b.1) Control panel: morphology of colonies rescued with wild type PouV proteins.
(b.2) Morphology of colonies transfected with Oct4 and Xlpou91 activator forms.
(b.3) Morphology of colonies transfected with repressor forms of Oct4 and Xlpou91.

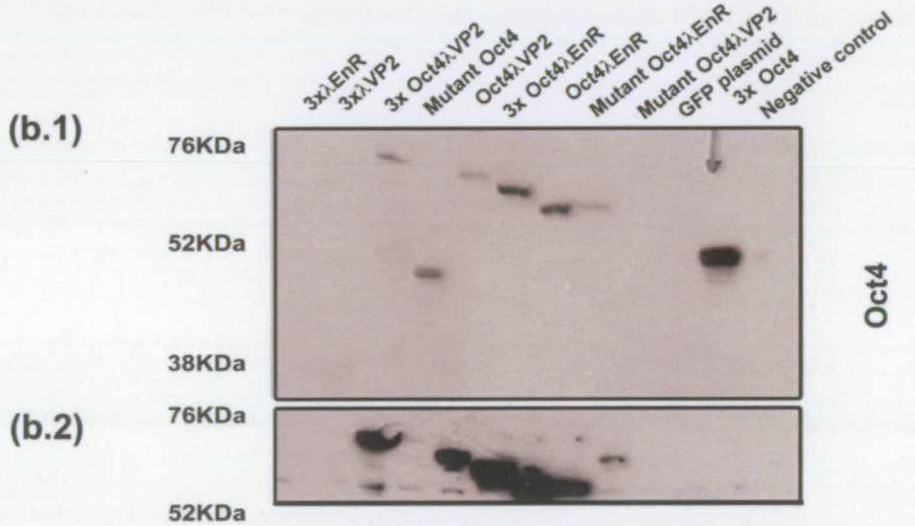
In order to exclude the possibility that the phenotypes generated by these fusion proteins were caused by non-specific activity of the reiterated regulatory domains employed, these domains were tested in the context of an Oct4 lacking the DNA binding ability. This lack of DNA binding ability is caused by a single amino acid change in the POU domain (from a Valine to Proline at position 267) (Niwa et al., 2002).

The absence of DNA binding activity in these new fusion proteins was confirmed in transient transfections with a set of Oct4 dependent reporter genes (Figure 3.2d and e). While these proteins are impaired in their capacity to activate or repress transcription from these reporters, they exhibited a very low level of residual activity. This activity is also observed when the DNA binding mutant Oct4 is tested on these reporters. Therefore, we conclude that this residual activity represents the recruitment of these proteins to DNA by other factors normally associated with Oct4, and is independent of the reiterated transcriptional regulatory units.

Moreover, to confirm that the reduced activity of these DNA binding mutants was not caused by a defective expression from the constructs, we examined the protein levels produced by these vectors when expressed transiently in HEK 293 cells (Figure 3.9). A western blot with an antibody to VP16 indicates that similar levels of proteins are expressed form both the DNA binding mutant, and the non-mutant activators.

The ability of these mutant proteins to rescue ES cells self-renewal was tested in the ZHBTc4 complementation assay. Both DNA binding mutant Oct4λVP2 and mutant Oct4λEnR generated AP negative differentiated colonies similar to those produced by the DNA binding mutant wild type Oct4 (Figure 3.8b1, b2, and b3), indicating that the rescue by the activator fusions required an intact Oct4 DNA binding domain, and was not a result of some unspecific effects contributed by the VP16 domain.


(a.1)
3xλEnR
3xλVP2
3x Oct4λVP2
Mutant Oct4
Oct4λVP2
3x Oct4λEnR
Oct4λEnR
Mutant Oct4λEnR
Mutant Oct4λVP2
GFP plasmid
3x Oct4
Negative control
76KDa
52KDa
38KDa
VP16
(a.2)
38KDa


(b.1)
3xλEnR
3xλVP2
3x Oct4λVP2
Mutant Oct4
Oct4λVP2
3x Oct4λEnR
Oct4λEnR
Mutant Oct4λEnR
Mutant Oct4λVP2
GFP plasmid
3x Oct4
Negative control
76KDa
52KDa
38KDa
Oct4
(b.2)
76KDa
52KDa

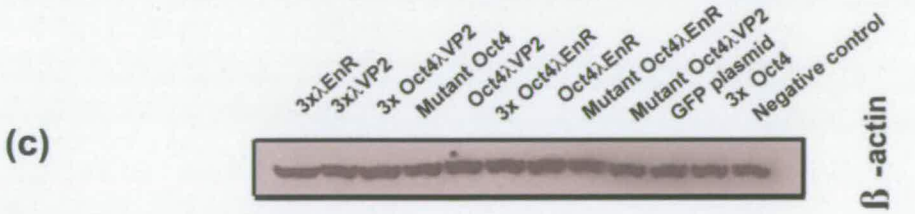

(c)
3xλEnR
3xλVP2
3x Oct4λVP2
Mutant Oct4
Oct4λVP2
3x Oct4λEnR
Oct4λEnR
Mutant Oct4λEnR
Mutant Oct4λVP2
GFP plasmid
3x Oct4
Negative control
β -actin

Figure 3.9-Western analysis of activator and repressor forms of Oct4 and their corresponding DNA binding mutants

Equal quantities (4ug) of CAG-Oct4λVP2, CAG-Oct4λEnR, or control constructs (CAG-DNA binding mutant Oct4λVP2, CAG-DNA binding mutant Oct4λEnR, and DNA binding mutant Oct4. CAG-3xFlag Oct4λVP2, CAG-3xFlag Oct4λEnR , 3xFlag Oct4, CAG-3xFlag λVP2, CAG-3xFlag λEnR) were transiently transfected into HEK293 cells.
24 hours after transfection, lysates were analysed by western blotting with anti VP16 and Oct4 antibodies.
Non-transfected cells were included as a negative control.

(a.1) Anti VP16 antibody at high exposure
(a.2) Anti VP16 antibody at low exposure
(b.1) Anti Oct4 antibody at high exposure
(b.2) Anti Oct4 antibody at low exposure
(c) ß-actin antibody

The predicted sizes for the proteins areas follow:
Oct4λVP2: 62.047 KDa
Oct4λEnR: 59.05 KDa
DNA binding mutant Oct4λVP2: 62.047 KDa
DNA binding mutant Oct4λEnR: 59.05 KDa
DNA binding mutant Oct4: 41.01 KDa
3xFlag Oct4λVP2: 64.89 KDa
3xFlag Oct4λEnR : 61.89 KDa
3xFlag Oct4: 43.85 KDa.
3xFlag λVP2: 23.88 KDa
3xFlag λEnR : 20.88 KDa

Interestingly, unlike the DNA binding mutant Oct4, no colonies were obtained when the DNA binding mutant activator or repressor domain fusions were introduced into cells already expressing *Oct4* (absence of Tc) (Figure 3.8b1, b2, and b3). As the differentiation induced by *Oct4* over-expression has been shown to be independent of DNA binding (Niwa et al., 2002), it has been suggested that high levels of Oct4 in a cell might "squelch" Oct4 mediated transcription by sequestration of Oct4 specific co-factors away from productive DNA bound complexes. A similar mechanism might be at work here and the reiteration of activation and repression domain modules could increase the affinity of non-DNA bound Oct4 for crucial co-factors required for ES cell growth.

3.2.4. Clonal cell lines can be derived from colonies rescued with PouV activator proteins and maintained in culture

As colonies rescued with Oct4λVP2 and Xlpou91λVP2 proteins exhibited an enhanced undifferentiated ES cell morphology, we wished to explore the basis for this phenotype by establishing cell lines in which these activators supported self-renewal in the absence of the Tc regulatable *Oct4* transgene expression.

Following the electroporation of vectors driving the expression of *Oct4*, *Oct4λVP2* and *Xlpou91λVP2* upstream of an IRES puromycin cassette in the ZHBTc4 cells in the presence of Tc, individual Puro resistant colonies were expanded in ES cell self-renewing conditions (under selection and in the presence of Tc). The same approach was used previously to derive cell lines expressing different PouV proteins in place of Oct4, and in no case was Oct4 expression detected in the expanded cell lines (Morrison and Brickman, 2006).

Figure 3.10 shows the morphology of clonal cell lines derived from colonies produced by Oct4λVP2 and Xlpou91λVP2 in absence or presence of tetracycline at passage 5. Oct4λVP2 and Xlpou91λVP2-supported cultures consist of highly compact undifferentiated cells with less spontaneous differentiation in the culture in comparison to the parental ZHBTc4 cell line, even when the Tc regulatable *Oct4* transgene is on.

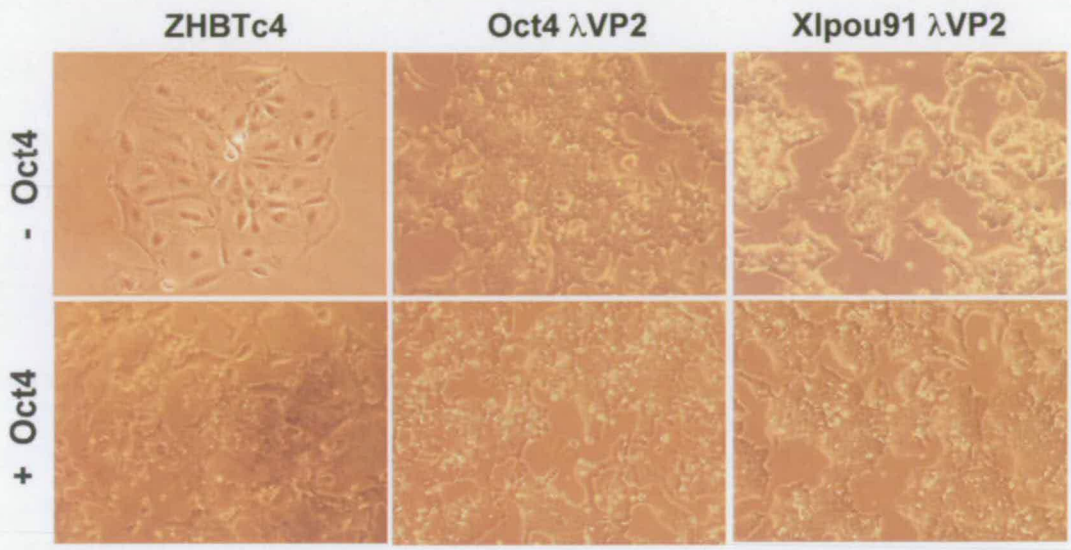


Figure 3.10-Derivation of clonal cell lines rescued with Oct4 and Xlpou91 activator forms

Morphology of Oct4λVP2 and Xlpou91λVP2 cell lines at p5 in the absence and the presence of Oct4 transgene expression.

Following the electroporation of Oct4λVP2 and Xlpou91λVP2 constructs in the ZHBTc4 cells, puromycin selection was applied for 9days in the presence (-Oct4) or absence (+Oct4) of Tetracycline. The rescued colonies were then picked and propagated.

3.2.5. Characterization of Oct4λVP2 cell line

3.2.5.1. Oct4λVP2 cell line supports normal self-renewal rate

To assess the relative efficiency of self-renewal of the activator rescued cell lines, we tested the capacity of single cells to self-renew and generate normal ES cell colonies when grown in the presence of serum and LIF.

As both Xlpou91λVP2, and Oct4λVP2 expressing cell lines showed similar phenotypes and behaviour, we chose to focus our analysis on Oct4λVP2 cell line. It offers the advantage of studying Oct4 activator function in its relevant physiological environment.

Cells were plated at clonal density and cultured for 6 days, and the resulting colonies stained for alkaline phosphatase (AP) activity. Colonies were scored as AP positive undifferentiated colonies, AP negative differentiated colonies or mixed colonies containing both undifferentiated and differentiated cells. Examples of these morphologies are shown in Fig 3.11a.

Based on this assessment of colony morphology, the activator expressing cell lines have a marginally increased rate of self-renewal over either those cells rescued by Oct4 or the parental ZHBTc4 cells (Figure 3.11b). These values represent average values from two experiments, each with two independent clones from the indicated cell line.

Interestingly, colonies formed by Oct4λVP2 expressing cell lines also exhibited distinct morphologies; they were smaller and less differentiated than normal Oct4-supported cell lines.

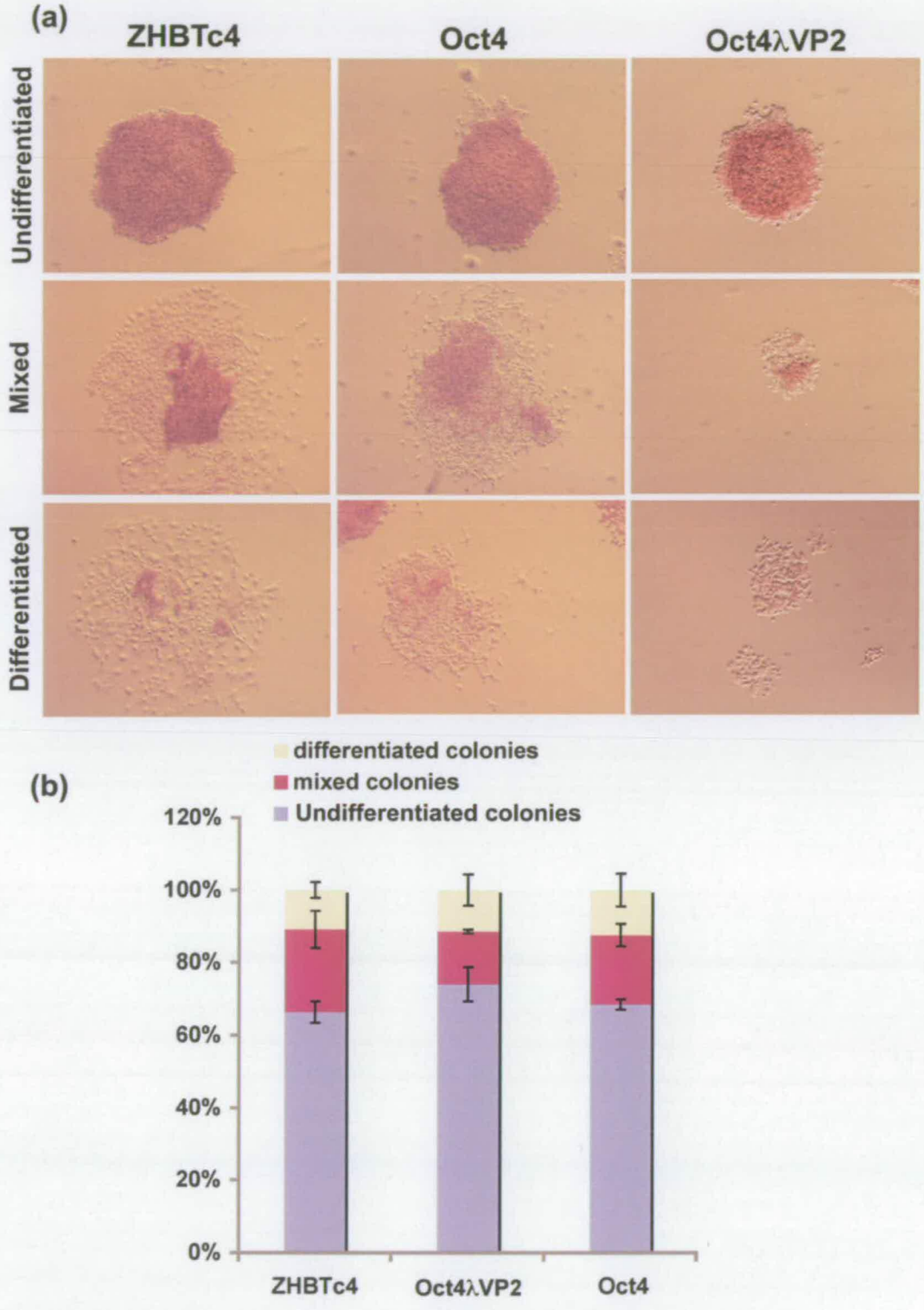


Figure 3.11-Self-renewal assay of Oct4 and Oct4λVP2 cell lines

The indicated cell lines were plated at clonal density (60 cells/cm2), and cultured for 6 days in self-renewing condition then stained for alkaline phosphatase (AP) activity (red)

(a) Examples of colony morphology present in each cell line. Colonies were classified into three categories: Uniformly AP positive undifferentiated colonies, mixed colonies containing AP positive and negative cells, and AP negative differentiated colonies.

(b) Percentage of previous colony categories in each cell line. Data represents average values from two experiments in which tow independent clones from each cell line were used.

3.2.5.2. Proliferation rate of the activator rescued cell lines

The observation that Oct4 activator supported cell lines formed smaller colonies than those of wild type Oct4 supported cells suggested that either growth or proliferation might be reduced in response to expression of Oct4λVP2. To investigate this possibility the proliferation rate of Oct4λVP2 supported cell lines was compared to the proliferation rate of Oct4 maintained cells, parental ZHBTc4 cells and wild type E14Tg2a ES cells. Proliferation was measured by quantifying the increases in viable cells over time via an MTS assay. At least two clones for each genotype were used to compile the data shown in Figure 3.12. The data indicates that there is little difference in the proliferation rates of the different cell lines. As a result, the difference in the colony size is not due to a slower growth rate, but rather some aspect of cell size or adhesion.

3.2.5.3. Molecular phenotype of activator (Oct4λVP2) rescued ES cells

As the colony size and morphologies observed in the activator supported cell lines were not a result of relative differences in proliferation, we examined expression of key pluripotency genes. Staining for Nanog, Sox2 and Oct4 revealed some interesting differences. An apparent subtle increase in Sox2 was observed in Oct4λVP2 supported cell lines alongside a slight reduction in overall levels *Oct4* transgene levels (Figure 3.13). However, the most striking observation was the relative increase in Nanog expression specific to the Oct4λVP2 cells. This was later confirmed by RT-PCR (Figure 3.23 a). At least three clones from each cell line were used for this analysis.

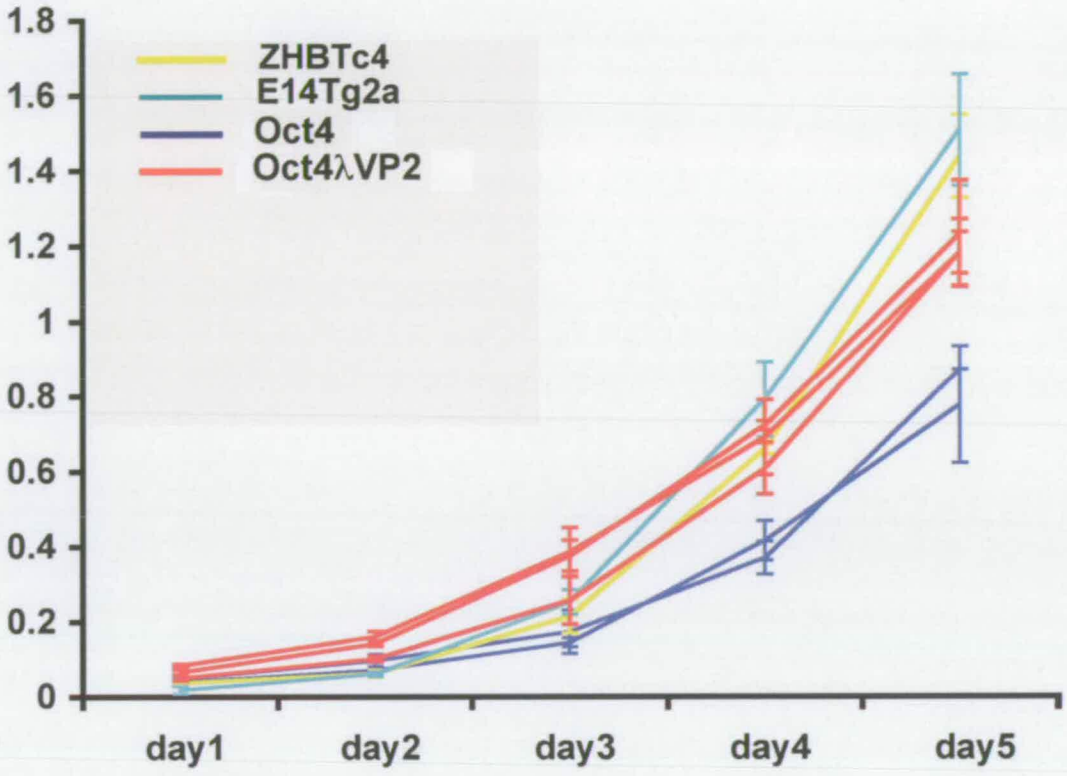


Figure 3.12-Oct4λVP2 rescued cell lines proliferate at a similar rate to wild type ES cells in Self-renewing condition

Cell growth was measured by the MTT assay and is plotted on the *y*-axis against time (days) on the *x*-axis. Data represents one of two experiments in which each data point was performed in triplicate.

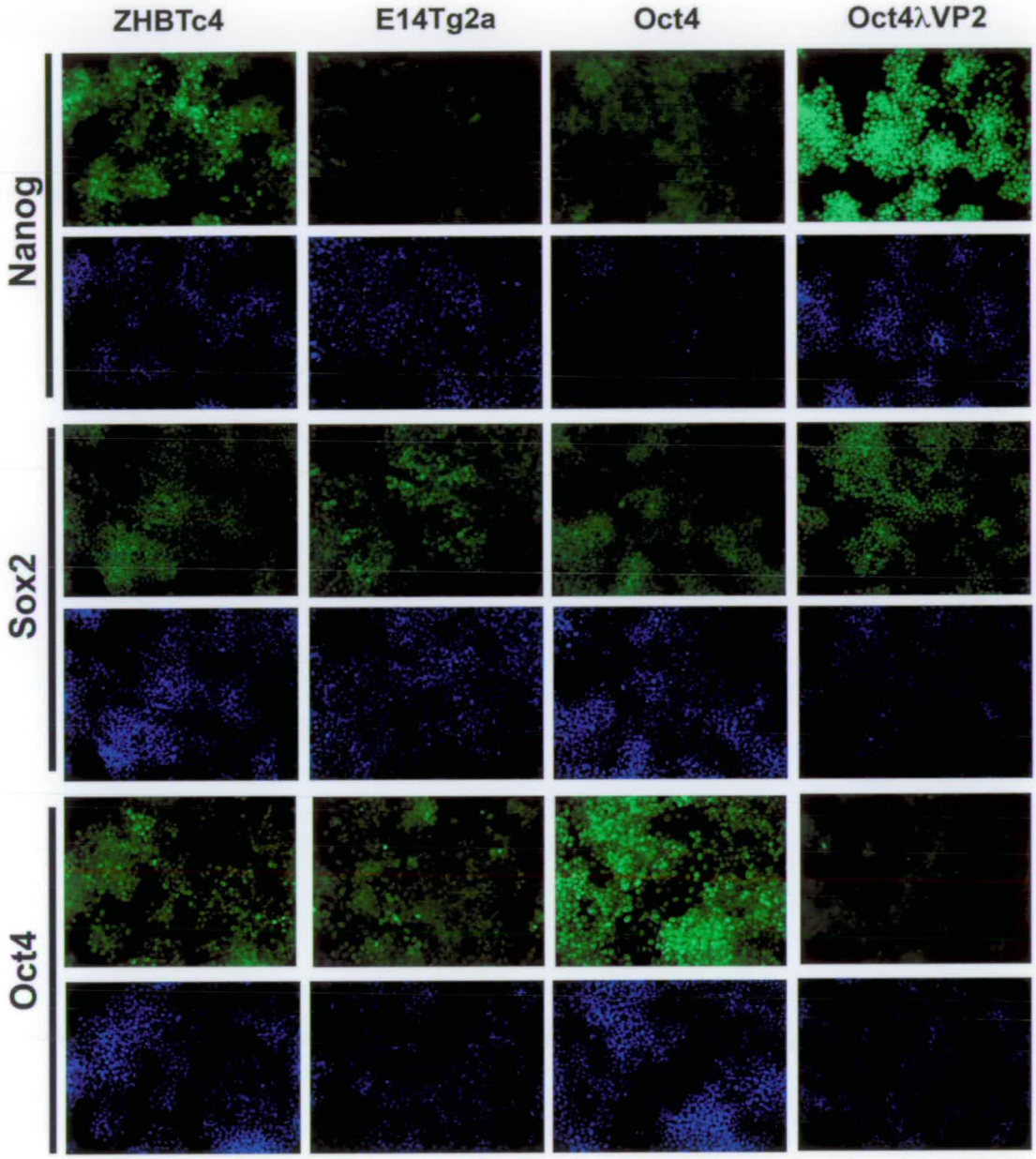


Figure 3.13-Immunocytochemistry of Oct4λVP2 cell line under self-renewing conditions

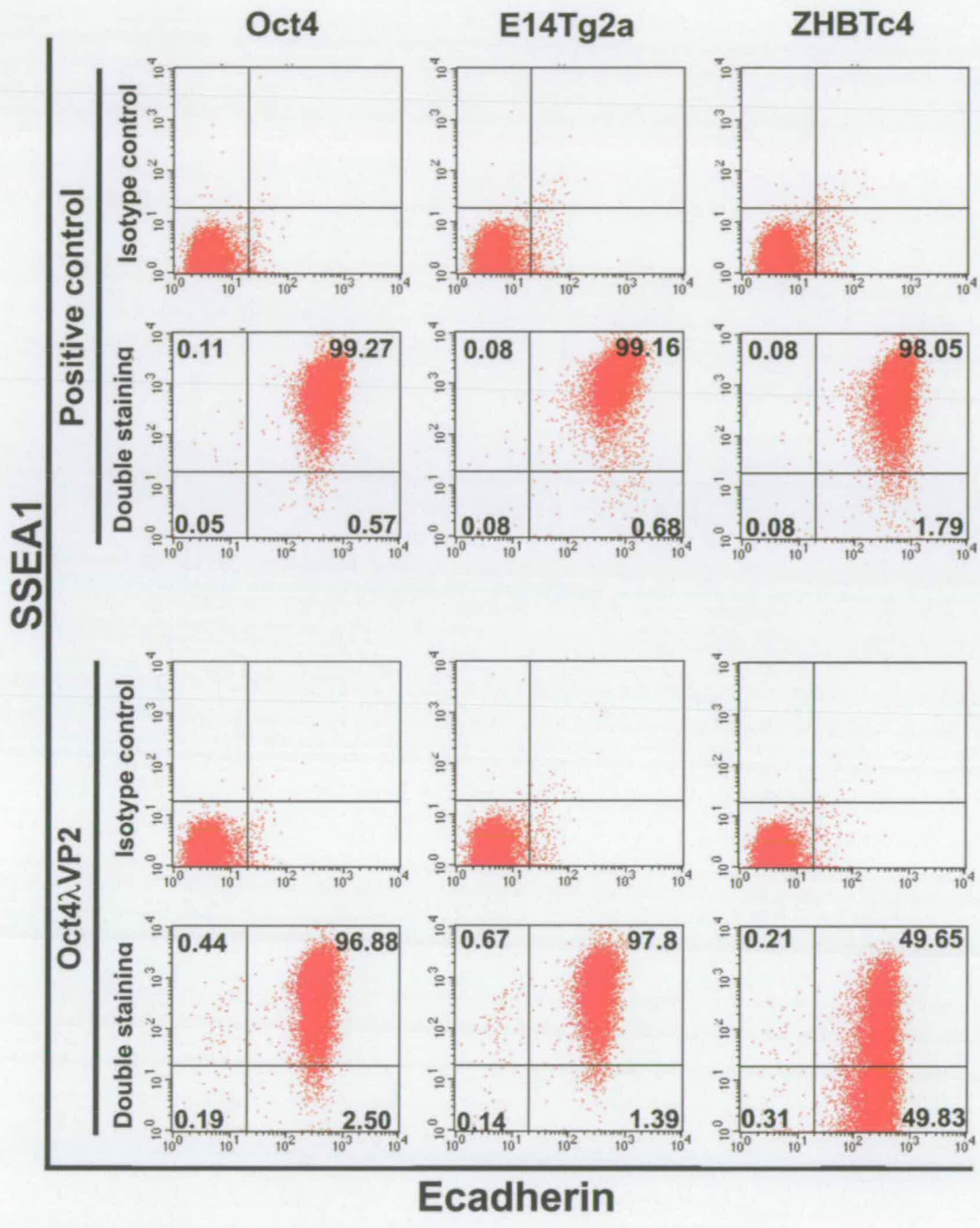
Three clones from each cell line (except ZHBTc4 and E14Tg2a one clone each) cultured under self-renewing conditions were fixed and immunostained for Oct4, Nanog and Sox2 proteins.
Green and Bright field images are shown for each cell line and each antibody.

We also examined cellular phenotypic variations present in typical ES cell cultures grown in LIF and serum. Flow cytometry was used to assess cell-based expression of a number of cell surface markers, known to characterize undifferentiated populations, in Oct4λVP2-supported ES cells. Cells were stained with antibodies specific to stage specific embryonic antigen-1 (SSEA-1), epithelial-calcium dependent adhesion molecules (E-cadherin), and platelet endothelial cell adhesion molecule-1 (PECAM-1 also known as CD31) (Boiani and Scholer, 2005; Cui et al., 2004).

While expression of both E-cadherin (Figure 3.14a) and PECAM-1 (Figure 3.14b) is in nearly all the cells of the Oct4λVP2 cell line, and is essentially indistinguishable in all clones, the expression of SSEA-1 does show some variation (Figure 3.14a). All three clones of Oct4λVP2 expressing cells have some reduction in the extent to which the population is SSEA-1 positive. Although this reduction is minimal in two clones, the third clone shows a significant reduction in about half the population. Interestingly, SSEA-1 heterogeneity in ES cells was reported previously (Cui et al, 2004), and while the extent of SSEA-1 expression can vary in different cell lines, there is little correlation between SSEA1 expression and the ability of ES cells to contribute to blastocysts.

3.2.5.4. Differentiation potential of the Oct4λVP2 cell line

The highly undifferentiated morphology and high levels of Nanog expression in Oct4λVP2 supported cell lines, in addition to the undifferentiated morphology of colonies generated by the PouV activator forms in the presence of Tc regulatable *Oct4* transgene expression suggest that these cells are resistant to differentiation induced by *Oct4* overexpression (Morrison and Brickman, 2006; Niwa et al., 2002). All these observations prompted us to examine the response of these cell lines to cues inducing differentiation.


Oct4
E14Tg2a
ZHBTc4
SSEA1
Positive control
Oct4λVP2
Isotype control
Double staining
0.11
99.27
0.05
0.57
0.08
99.16
0.08
0.68
0.08
98.05
0.08
1.79
0.44
96.88
0.19
2.50
0.67
97.8
0.14
1.39
0.21
49.65
0.31
49.83
Ecadherin

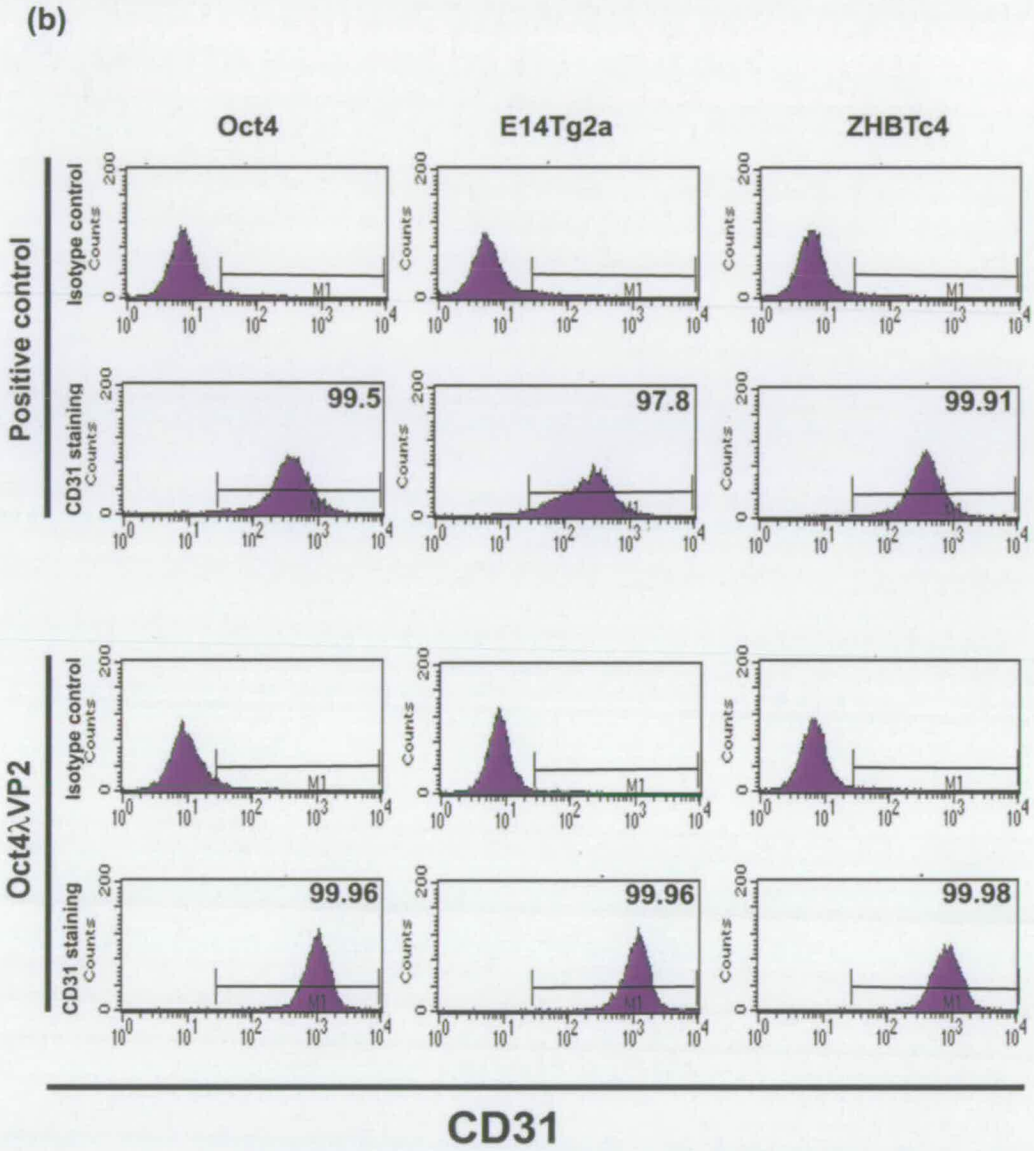


Figure 3.14-Oct4λVP2 cell line expresses ES cell associated cell surface markers under self-renewing conditions

Three clones at least from Oct4λVP2 and Oct4 cell lines growing in self-renewing conditions were used for FACs analysis (only one clone shown for Oct4 cell line). ZHBTc4 and E14tg2a cell lines were used as positive controls.

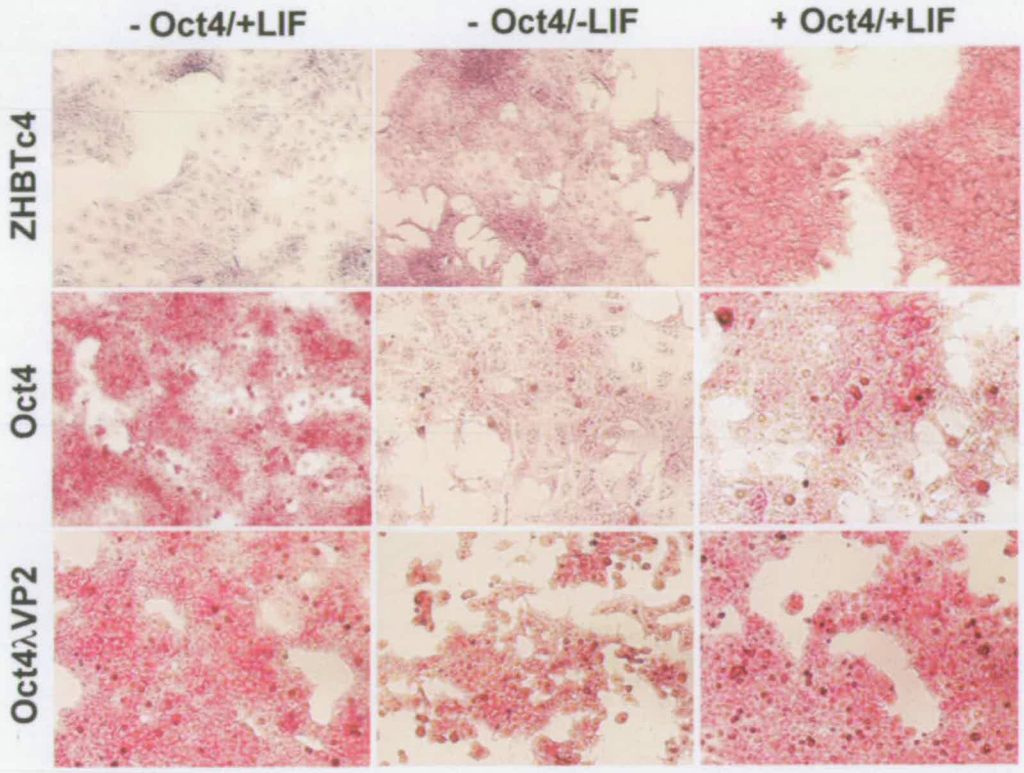
(a) FACs analysis for SSEA-1 and Ecadherin double staining is shown for each cell line.
(b) FACs analysis for CD31 staining for each cell lines.
Numbers are the percentages of cells in each of the indicated gates

We tested the consequences of re-expression of Tc regulatable *Oct4* transgene, as well as LIF withdrawal on the phenotype of activator supported cell lines.

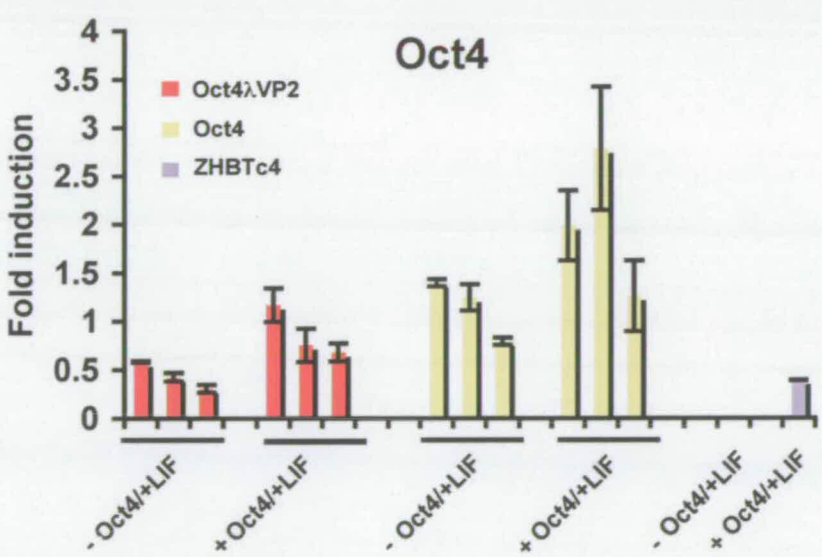
Figure 3.15a1 shows that unlike Oct4, and ZHBTc4 cell lines, Oct4λVP2 cells remain undifferentiated and positive for AP staining when *Oct4* transgene is reactivated. Although, Oct4 levels increased upon reactivation of the Tc regulatable *Oct4* transgene in both Oct4 and Oct4λVP2-supported cell lines (Figure 3.15 a2), the total Oct4 levels in Oct4λVP2 cell line after the reactivation of *Oct4* transgene are comparable to that of Oct4 cell line growing in self-renewal conditions. Suggesting that the total Oct4 levels in Oct4λVP2 cell line after the reactivation of Tc regulatable *Oct4* transgene might have not reached the threshold level required to induce ES cell differentiation.

While the response of activator supported cell lines to Oct4 overexpression might not be that surprising, Figure 3.15a also shows that these lines are resistant to differentiation induced by LIF withdrawal. Normally ES cell cultures require LIF to maintain the pluripotent state, and in its absence, ES cells normally differentiate (Niwa et al., 2000; Smith et al., 1988; Williams et al., 1988).
As predicted, when cultured for five days in the absence of LIF, both the ZHBTc4 cells and Oct4 rescued cell lines adopt a differentiated morphology and loose the AP expression (Figure 3.15 a1). However, Oct4λVP2 supported ES cells appear not to differentiate, instead they retain their AP positive undifferentiated morphology, and expand at reduced rates. Similar results were obtained when self-renewal in the absence of LIF was assayed at a clonal level. *Oct4λVP2*-expressing cell lines generated tiny AP positive colonies when plated at clonal density in the absence of LIF (Figure 3.15b1). When these colonies were allowed to grow for another ten days, a significant proportion of them formed normal size undifferentiated colonies (30%) (Figure 3.15b2, and b3). Moreover, the colonies scored as mixed in the activator supported cell lines, were scored based on colony size rather than phenotype. These results suggest that Oct4λVP2 cell line does not differentiate upon LIF withdrawal, instead, they self-renew at a slower rate showing signs of LIF-independent self-renewal.

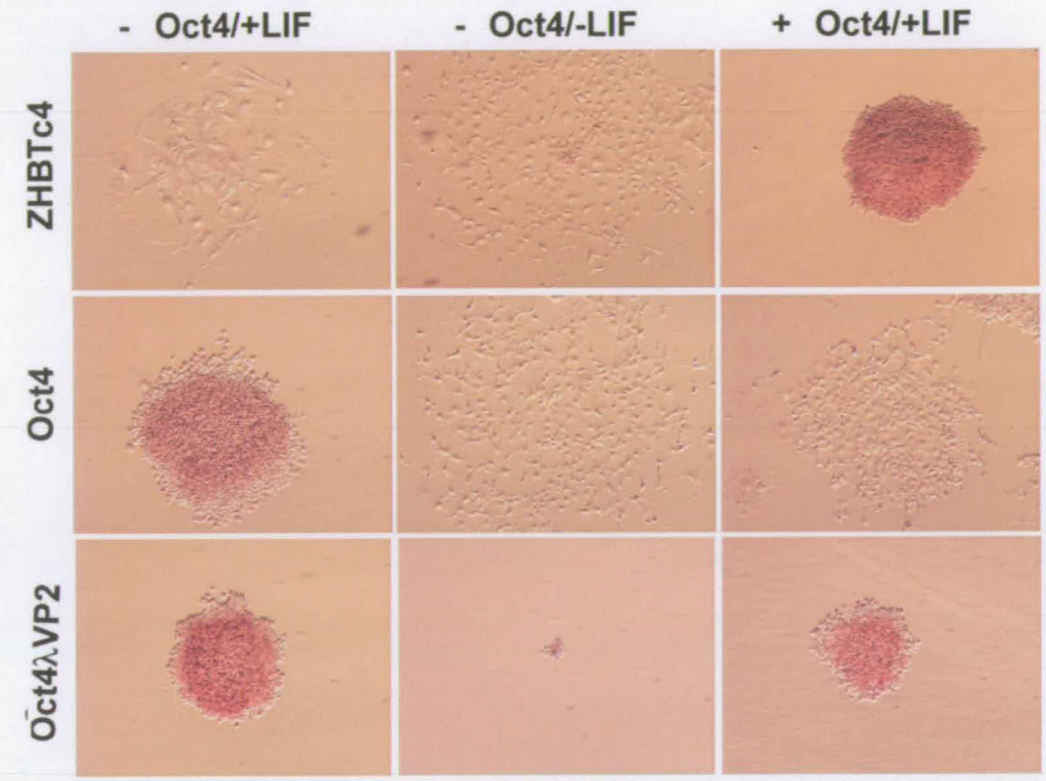
(a.1)


- Oct4/+LIF
- Oct4/-LIF
+ Oct4/+LIF
ZHBTc4
Oct4
Oct4λVP2

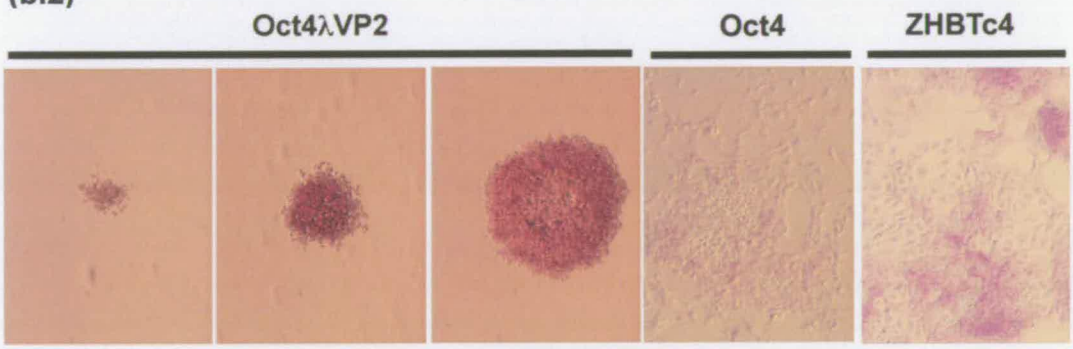
(a.2)


Oct4
Oct4λVP2
Oct4
ZHBTc4
Fold induction
4
3.5
3
2.5
2
1.5
1
0.5
0
- Oct4/+LIF
+ Oct4/+LIF
- Oct4/+LIF
+ Oct4/+LIF
- Oct4/+LIF
+ Oct4/+LIF

(b.1)


- Oct4/+LIF
- Oct4/-LIF
+ Oct4/+LIF
ZHBTc4
Oct4
Oct4λVP2

(b.2)


Oct4λVP2
Oct4
ZHBTc4

(b.3)

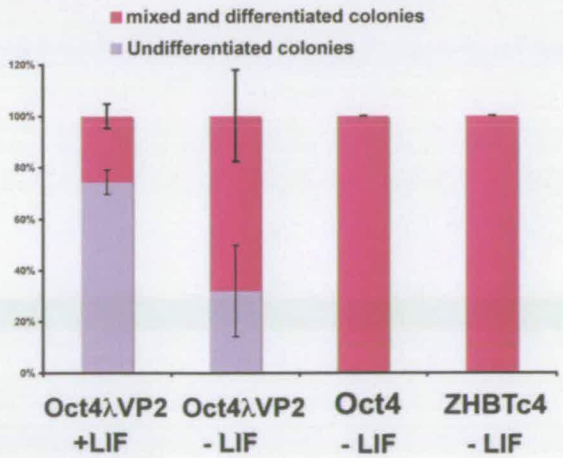

mixed and differentiated colonies
Undifferentiated colonies
120%
100%
80%
60%
40%
20%
0%
Oct4λVP2 +LIF
Oct4λVP2 - LIF
Oct4 - LIF
ZHBTc4 - LIF

Figure 3.15-Oct4λVP2 cell line does not differentiate upon LIF withdrawal or in Oct4 overexpression conditions

(a.1) Monolayer differentiation of Oct4λVP2 cell line: cells were plated in the absence of LIF, or Oct4 overexpression conditions for 5 days. As well as in self-renewing conditions as a control.
Oct4 overexpression condition is achieved by removal of tetracycline from the medium leading to the re-expression of the Tc regulatable Oct4 transgene.
Three clones for each cell line except ZHBTc4 were used and stained for AP activity (red). Photos of representative clones are shown.

-Oct4/+LIF represents self-renewing condition.
-Oct4/-LIF represents LIF withdrawal condition.
+Oct4/+LIF represents mOct4 overexpression condition.

(a.2) qRT-PCR analysis of Oct4 mRNA levels. RNA was collected at day 5 of the differentiation experiments. Levels of Oct4 expression are relative to TBP (Tata binding protein) level.

(b.1) Clonal density differentiation of Oct4λVP2 cell line: cells were plated at clonal density in the absence of LIF or Oct4 overexpression condition for 5 days as well as self-renewing conditions as a control.
The same three clones for each cell line were used and stained for AP activity (red). Photos of representative clones are shown.

(b.2) Morphology of Oct4λVP2 cells growing in the absence of LIF for 15 days cells from clonal density differentiation were allowed to grow in culture in the absence of LIF, and then stained for AP activity.

3.2.6. Investigating the response of Oct4λVP2 cell line to LIF withdrawal

To understand the molecular basis for the response of Oct4 activator-supported ES cells to LIF withdrawal, and investigate its capacity for LIF independent self-renewal, we carried out microarray gene expression analysis on RNA derived from a set of PouV rescued cell lines including Oct4λVP2, Oct4 and the parental ZHBTc4 cells in response to LIF withdrawal. An outline of the experimental plan is given in Figure 3.16. The experiment was designed to compare gene expression under standard self-renewing conditions and then to follow those gene expression changes into differentiation. Cells were plated at low density under self-renewing conditions for 24 hours, at this point RNA time point was taken from each cell line. The culture medium was changed to an identical one without LIF and two additional time points were taken at 48 and 120 hours after LIF withdrawal.

Two independent clones from each cell line were used as both biological and technical replicates. Samples were hybridised to the NIA Mouse 44K Microarray v2.1 (Carter et al., 2005), and gene expression analysed in collaboration with Minoru Ko and Alexei Sharov at the NIH.

ANOVA analysis was performed by A. Sharov using standard statistical conditions (FDR<0.05, 2-fold expression levels change) to unveil genes with changes in expression levels between the samples and controls.

Statistical analysis was done using the NIA Array Analysis software (Sharov et al., 2005).

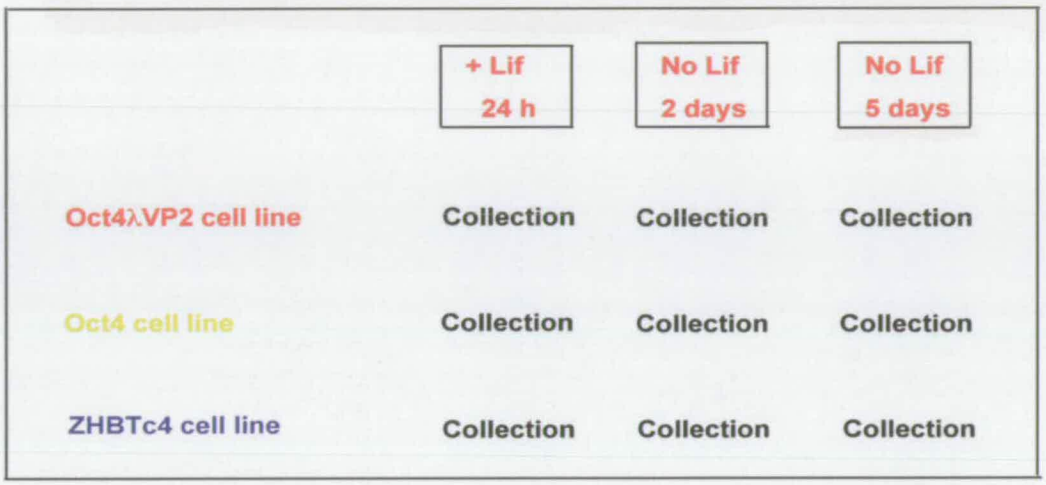


Figure 3.16-Schematic representation of the micro-array experiment design

Cells were plated at low density (3000 cells/cm2) under self renewing conditions for 24 hours; at that point RNA from all the cell lines was collected. The culture medium was then changed to medium without LIF. After 48 hours of LIF withdrawal, RNA was collected from all cell lines for the second point. At this stage, the cells were fed again with medium without LIF and left in culture for another 3 days before RNA was collected for the final point.

Two clones from each cell lines were used as biological replicate in the whole experiment.

3.2.6.1. Genes responded to LIF withdrawal differently in Oct4λVP2 cell line and Oct4/ZHBTc4 cell lines

Gene expression analysis was performed to unveil genes that respond to LIF withdrawal in Oct4-supported and ZHBTc4 cell lines and study their behaviour in Oct4λVP2-supported cell line after LIF withdrawal.

Figure 3.17a shows that the majority of genes upregulated or downregulated by more than two fold upon LIF withdrawal in either Oct4 rescued or parental ZHBTc4 cell lines did not change upon LIF withdrawal in Oct4λVP2-supported cell lines. Moreover, the genes that did change in the activator support lines could not be associated with a particular trend in either control cell lines (Figure 3.17b). This difference in the gene expression response of the Oct4λVP2 cell lines correlates with the distinct undifferentiated, but slow growing phenotype observed in these cells when cultured in the absence of LIF, and suggests that this cell line is distinct at the molecular level from the two other cell line.

3.2.6.2. The majority of genes regulated by Oct4 are regulated in the same direction by Oct4λVP2

Previous work in the laboratory of our collaborator Minoru Ko has used the ZHBTc4 cells to define a set of Oct4 responsive genes and targets (Matoba et al., 2006; Sharov et al., 2008). We have used these data sets to address the question of how the genes up or down regulated by Oc4λVP2 compared to the genes defined as responding to Oct4 withdrawal in the ZHBTc4 cells. Figure 3.18 shows a plot of the total set of genes responding to Oct4 depletion in the ZHBTc4 cells compared to the enhanced effect of Oct4λVP2 on gene expression. A positive value in Figure 3.18 means that the gene is activated, and a negative value means the gene is suppressed.

Not surprisingly, inspection of Figure 3.18 indicates that the majority of genes positively regulated by Oct4 are also positively regulated by Oct4λVP2. However, the plot also suggests that the majority of Oct4 negatively regulated genes are also negatively regulated by Oct4λVP2. This is particularly striking, as it would imply that repression by Oct4 is predominantly indirect.

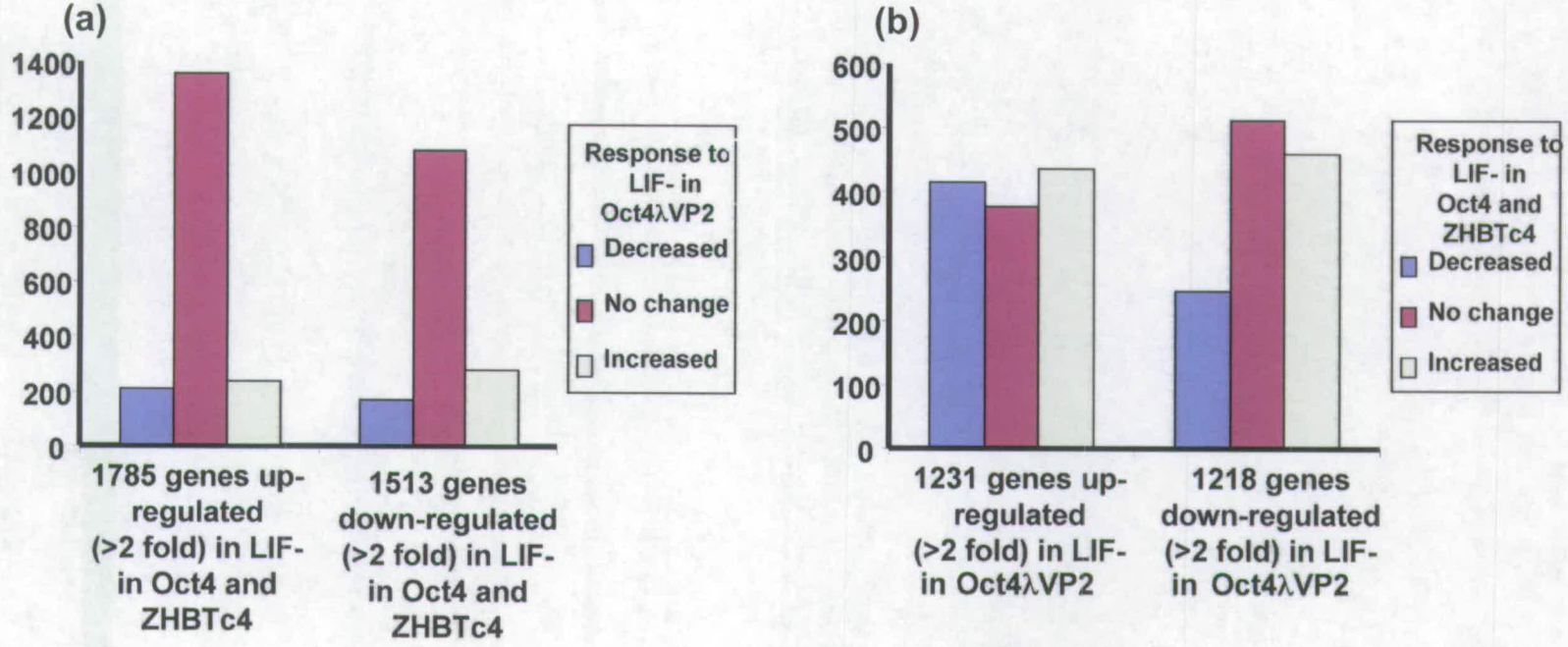


Figure 3.17-Genes responded to LIF withdrawal differently in Oct4λVP2 cell line and Oct4/ZHBTc4 cell lines

(a) Genes upregulated or downregulated upon LIF withdrawal in Oct4 and ZHBTc4 cell lines respond differently to LIF withdrawal in Oct4λVP2 cell line.
(b) Genes upregulated or downregulated upon LIF withdrawal in Oct4λVP2 cell line respond differently to LIF withdrawal in Oct4 and ZHBTc4 cell lines.

The Y-axis represents the number of genes in each subgroup.

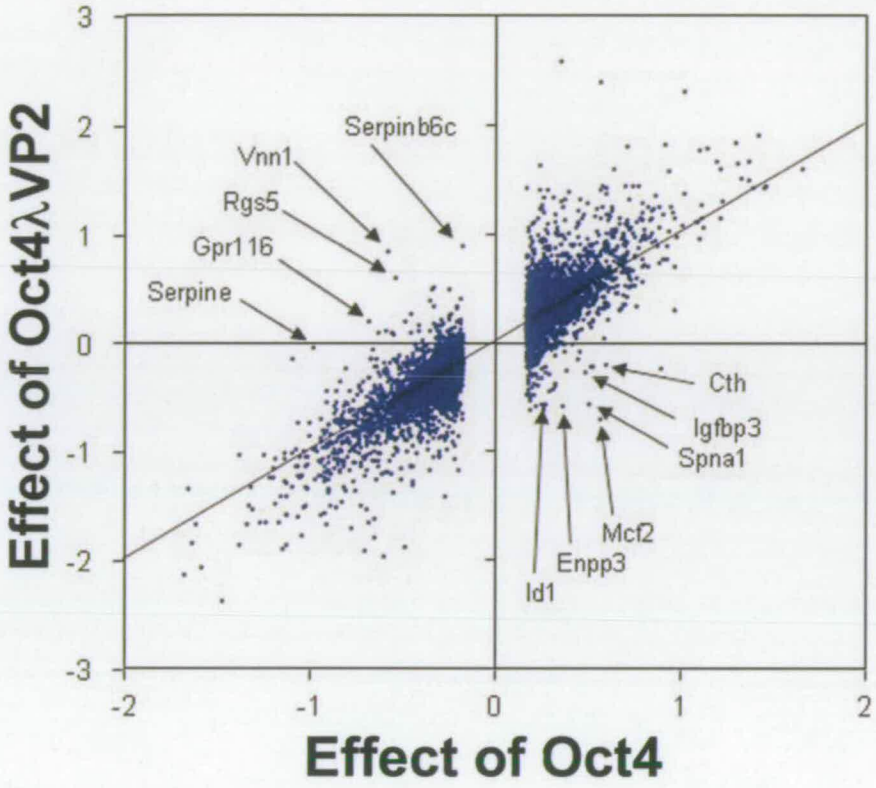


Figure 3.18-Most of the genes regulated by wild type Oct4 are regulated by its activator form (Oct4λVP2) in the same manner

The plot shows Oct4λVP2 effect versus mOct4 effect on gene expression.
Positive value means that the gene is activated and, negative value means the gene is suppressed.
Most genes that are activated by Oct4 are also activated by Oct4λVP2, and most genes that are suppressed by Oct4 are suppressed by Oct4λVP2.

It seems that both activation and suppression are in average stronger for Oct4λVP2 than for Oct4. However there are genes that are affected in opposite way by Oct4 and Oct4λVP2 (some of them are named on the figure).

To further investigate whether we could detect any evidence for the existence of true Oct4 repressed targets we examined the complete set of genes inferred by Sharov and colleagues as "directly suppressed" by Oct4 individually (Sharov et al., 2008). We reasoned that if Oct4 activity is involved in direct lineage specific repression, then the binding of Oct4λVP2 to these promoters would be expected to at least de-repress their transcription and we should observe some evidence of this in our microarray data. Table 3.1 lists the complete set of 52 genes inferred as directly repressed Oct4 targets as defined by Sharov et al 2008. Inspection of the behavior of these genes in our microarray indicates that 83% (43 genes) of them are repressed by Oct4λVP2, which supports the notion that the majority of these targets are not genuine direct Oct4 targets, and are instead indirectly repressed targets that are regulated by other aspects of the network downstream of Oct4.

As the activity of certain Oct4 repressed promoters such as *Cdx2* are vital for the maintenance of ES cell phenotypes, we further validated the conclusion from our microarray data by RT-qPCR on RNA obtained from two different clones from each cell line (Figure 3.19). All eight targets examined behaved as in the microarray analysis with the expression of some actually being lower in the Oct4λVP2 supported cell lines than in Oct4 rescued or the parental ZHBTc4 cells. Taken together our data supports the notion that Oct4 is a positive regulator of a network that silences gene expression associated with embryonic differentiation.

3.2.6.3. Difference in gene expression between Oct4λVP2 and to Oct4 and ZHBTc4 cell lines

To identify differentially regulated genes that might be associated with the properties of Oct4λVP2-supported cells, we compared gene expression in this cell line to those supported by wild type Oct4 (Oct4 rescued, and ZHBTc4).

Gene ontology (GO) annotations of upregulated and downregulated genes in Oct4λVP2 cell line with more than 2 fold change are listed in Table 3.2 and 3.3 and attached as appendices.

Table 3.1- List of Oct4 repressed targets

List of Oct4 repressed targets concluded from study by Sharov et al 2008. The microarray analysis of Oct4λVP2 cell line showed that most of the genes inferred to be repressed by Oct4 are still repressed in the mOct4λVP2 cell line except from genes in red colour.

Symbol	Gene name
Sdf2	stromal cell derived factor 2
Msx2	homeobox, msh-like 2
Fgfr1	fibroblast growth factor receptor 1
Cdx2	caudal type homeo box 2
Otx2	orthodenticle homolog 2 (Drosophila)
Dkk1	dickkopf homolog 1 (Xenopus laevis)
Tec	tec protein tyrosine kinase
Akap2	A kinase (PRKA) anchor protein 2
Rb1	retinoblastoma 1
Ches1	checkpoint suppressor 1
Ak3	adenylate kinase 3
Sptlc2	serine palmitoyltransferase, long chain base subunit 2
Arrb1	arrestin, beta 1
Fgf5	fibroblast growth factor 5
Eomes	eomesodermin homolog (Xenopus laevis)
Egln3	EGL nine homolog 3 (C. elegans)
Smarca2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2
1700034H14Rik	RIKEN cDNA 1700034H14 gene
Hsp90aa1	heat shock protein 90, alpha (cytosolic), class A member 1
Ddit4	DNA-damage-inducible transcript 4
2610042L04Rik	RIKEN cDNA 2610042L04 gene
9030025P20Rik	RIKEN cDNA 9030025P20 gene
Ss18	synovial sarcoma translocation, Chromosome 18
Higd2a	HIG1 domain family, member 2A
Dnajb14	DnaJ (Hsp40) homolog, subfamily B, member 14
AI450540	expressed sequence AI450540
LOC236749	similar to Testin (TESS
Cfdp1	craniofacial development protein 1
Ube2f	ubiquitin-conjugating enzyme E2F (putative)
Phactr1	phosphatase and actin regulator 1
Kctd10	potassium channel tetramerisation domain containing 10
Dmrt1	doublesex and mab-3 related transcription factor 1
Rpusd4	RNA pseudouridylate synthase domain containing 4
Tbx15	T-box 15
Fnbp1	formin binding protein 1
Son	Son DNA binding protein
2810003C17Rik	RIKEN cDNA 2810003C17 gene
Ccdc68	coiled-coil domain containing 68

Symbol	Gene name
Gpc4	glypican 4
Sema6a	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6A
Fat1	FAT tumor suppressor homolog 1 (Drosophila)
Qk	quaking
4933403G14Rik	RIKEN cDNA 4933403G14 gene
Fgfr2	fibroblast growth factor receptor 2
Uhrf2	ubiquitin-like, containing PHD and RING finger domains 2
Epb4.9	erythrocyte protein band 4.9
Lgmn	legumain
Prtg	protogenin homolog (Gallus gallus)
Utrn	utrophin
Cidea	cell death-inducing DNA fragmentation factor, alpha subunit-like effector A
Scap	SREBF chaperone
Gadd45g	growth arrest and DNA-damage-inducible 45 gamma

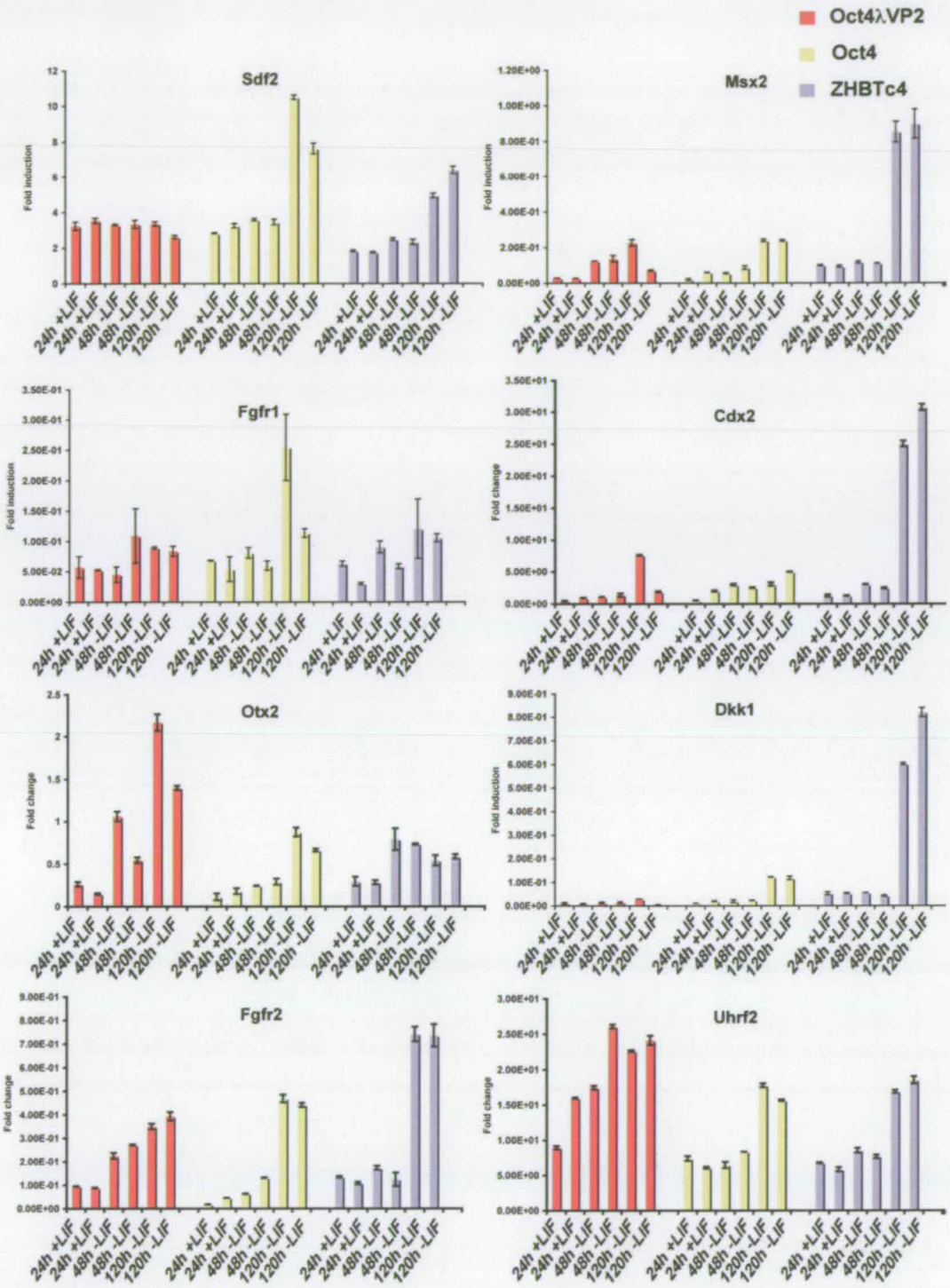


Figure 3.19-Most genes inferred to be repressed by Oct4 are still repressed in the Oct4λVP2 cell line

Microarray validation by qRT-PCR of few genes from the list of Oct4 repressed targets (Sharov et al 2008).

Total RNA was collected during the LIF withdrawal time course as explained previously in the microarray experiment design. Expression levels of all analysed genes were normalised to TBP levels for each sample.

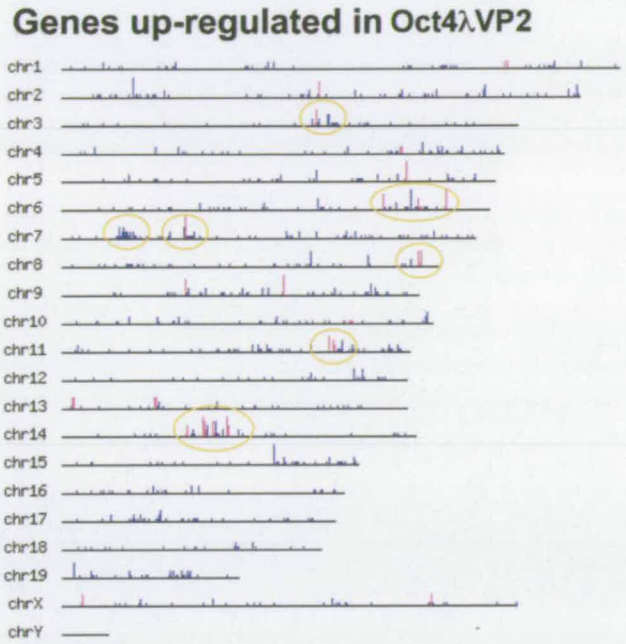
While the upregulated genes in Oct4λVP2 cells are enriched for functions related to plasma membrane, cell adhesion, extracellular matrix, and immune response, the downregulated genes are enriched for transcription factor activity, nervous system development, neurogenesis, growth factor activity, and cytokine activity.

The mapping of these gene expression changes onto the genome reveals striking clustering. In particular, genes upregulated in the activator supported cell lines appear to cluster in strong clusters on chromosomes 3, 5, 6, 7, 8, 11 and 14 (Figure 3.20a and b). A complete list of the 15 strongest clusters of upregulated genes in Oct4λVP2-supported cells is attached in the appendix (Table 3.4).

Interestingly, Oct4λVP2 cell line overexpresses many germ line and oocyte-specific genes. For example, *Oas* (cluster 3), and *Omt2b* (cluster 8) are known oocyte markers and *Dppa3* also known as *Stella* (cluster 5), and *Gsg1* (cluster 6) are associated with germ cell specification.

Surprisingly, this analysis also revealed that Oct4λVP2 cell line expressed high levels of the following neural genes: *Pax6*, *En1*, *NtrK1*, *Ntn1*, *Notch4*, *Nrxn3*, *En2*, *Gli2*, and *Hhip*. These genes are all upregulated in cells grown in LIF and serum and their expression is maintained during LIF withdrawal (Figure 3.21a). These observations were particularly intriguing as PouV proteins in lower vertebrates are expressed in the early neural plate and later in the isthmus region (Belting et al., 2001). As it had been previously shown that ectopic *Oct4* expression in the ZHBTc6 cells enhances neural gene expression (Niwa et al., 2002; Niwa et al., 2000), we tested whether activator supported lines undergo neural differentiation. In our hands, neither the ZHBTc4 cells them-selves nor any of our Oct4λVP2 or Oct4 supported lines underwent robust differentiation to morphologically normal *Tuj1* positive neurons (Figure 3.21 b). We believe this is likely due to the inability of these cell lines to downregulate Oct4 and undergo terminal differentiation. However, when our lines were compared to each other, the Oct4λVP2-supported cells consistently expressed higher levels of both *Nestin* and *Tuj1*, although morphologically they do not appear as neurons.

(a)



(b)

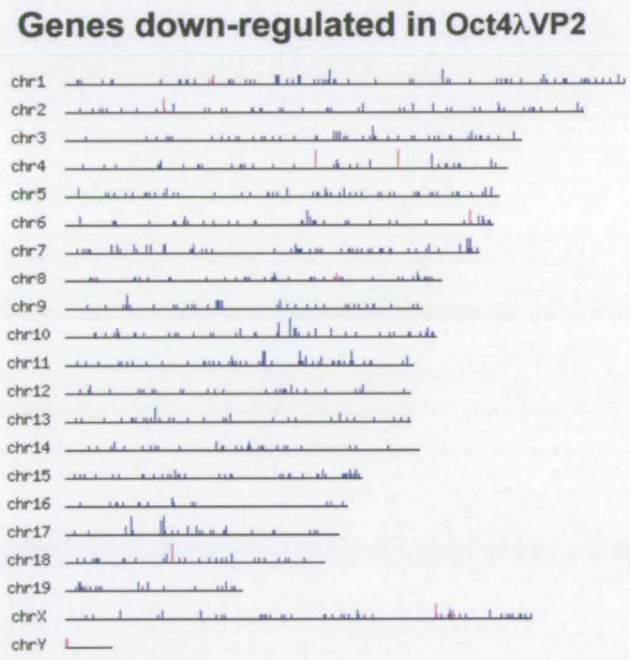
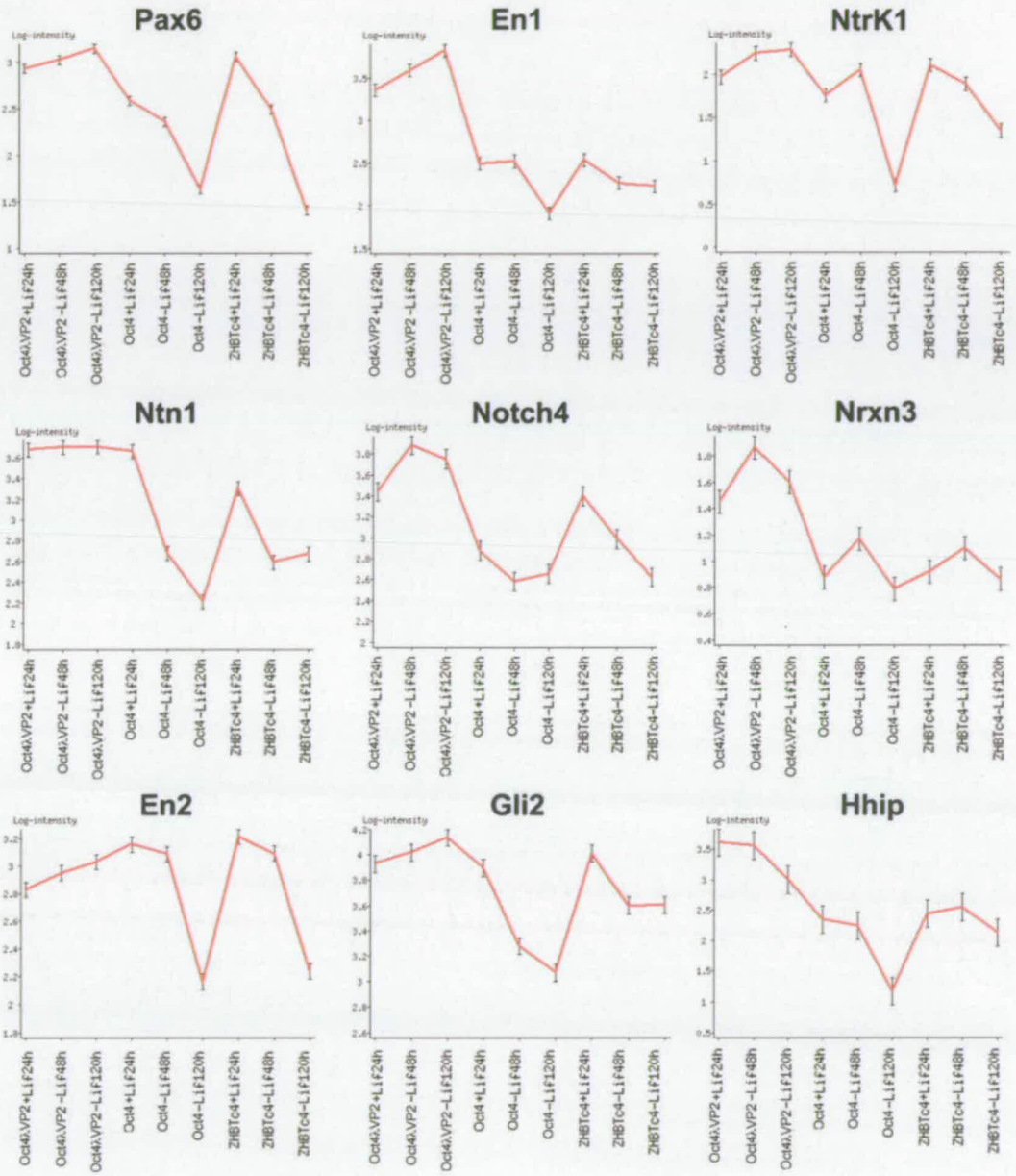


Figure 3.20-Genome location of genes affected by Oct4λVP2

(a) Genome location of clusters of upregulated genes in Oct4λVP2 cell line in comparison to Oct4 and ZHBTc4 cell lines.
(b) Genome location of genes downregulated in Oct4λVP2 cell line in comparison to Oct4 and ZHBTc4 cell lines.
The graphs were generated using NIA Mouse Gene Index (Sharov AA, et al 2005). The height of bars shows the number of genes per 1Mb of sequence. Magenta colour means that the number of upregulated or down regulated genes in the Oct4λVP2 cell line is significantly over-represented in this 1Mb chromosome region (FDR<0.1).
Blue colour means that the number of genes is not significantly over-represented.

(a)


Pax6
En1
NtrK1
Ntn1
Notch4
Nrxn3
En2
Gli2
Hhip
Log-intensity
Oct4\VP2+Lif24h
Oct4\VP2-Lif48h
Oct4\VP2-Lif120h
Oct4+Lif24h
Oct4-Lif48h
Oct4-Lif120h
ZHBTc4+Lif24h
ZHBTc4-Lif48h
ZHBTc4-Lif120h

(b)

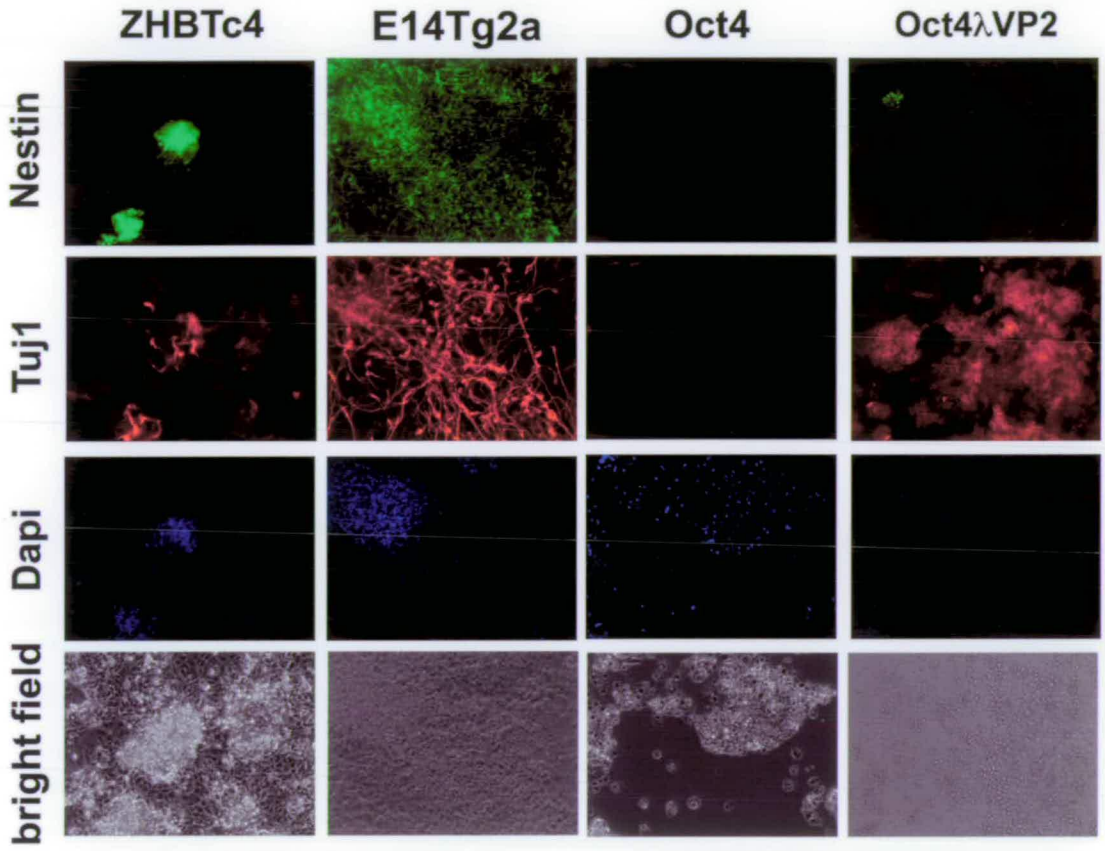


Figure 3.21-Oct4λVP2 cell line express high levels of neural genes

(a) Neural genes microarray plots: the plots are comparing log intensity values for each gene probe among the three cell lines at the three time points of the LIF withdrawal experiment. The error bars represents standard deviation between two biological replicates (two independent clones).

(b) *In Vitro* neural monolayer differentiation: three clones from each cell line (except ZHBTc4 and E14Tg2a one clone each) cultured under neural differentiation conditions for 6 days were fixed and immunostained for Nestin and Tuj1 proteins.
Nuclear staining (Dapi) and Bright field images are shown for each cell line.

While the neural signature comes clearly out of this dataset, only *Nanog* among known key pluripotency genes is overexpressed in the Oc4λVP2 cell line (Figure 3.22). However, other ES cell associated transcripts (ECATs) (Takahashi and Yamanaka, 2006): *Gdf3, Dppa2, Fbxo15, Dppa3, ERas*, and *Sox15* are also upregulated in the Oct4λVP2 cell line. It will be interesting to analyse the role of these genes in maintaining pluripotency, and their upstream regulators (Oct4 only targets, Nanog only targets, or target of both Oct4, and Nanog).

Upregulation of *Nanog* in the Oct4λVP2 cell line suggests it might play an important role in its response to LIF withdrawal, especially that *Nanog* overexpression is known to generate LIF-independent self-renewal (Chambers et al., 2003).

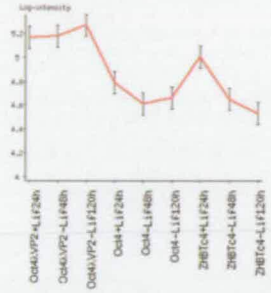
3.2.6.4. Gene expression analysis of Oct4λVP2 cells during LIF withdrawal time course

While *Nanog* was the only known key pluripotency gene upregulated in the undifferentiated activator supported cell lines, expression of other pluripotency genes was maintained during the LIF withdrawal time course (Figure 3.22). To explore these changes in gene expression more carefully, we validated these findings using real time qRT-PCR analysis of gene expression levels of the pluripotency genes *Oct4*, *Nanog*, *Klf4*, *Sox2*, *Rex1* and *Fgf4* during LIF withdrawal time course. While *Nanog* is noticeably upregulated in the undifferentiated Oct4λVP2 supported cell lines, the other pluripotency markers are all expressed at the same levels as in the control cell lines (Oct4 supported and ZHBTc4 parental cell lines). However, during LIF withdrawal, expression levels of all these pluripotency genes, except *Fgf4*, are maintained in Oct4λVP2 supported cells (Figure 3.23a).

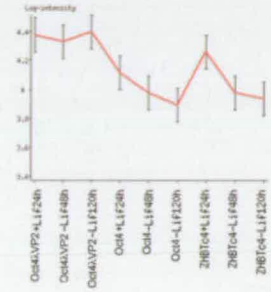
Interestingly, while most differentiation specific genes behaved in a similar manner during the LIF withdrawal time course in both Oct4λVP2 supported cells and control cell lines, *Gata4* and specially *Gata6* gene expression levels are significantly downregulated in the undifferentiated Oct4λVP2 cells and fail to increase upon LIF withdrawal. This suggests that the activator supported cell lines fail to from primitive endoderm upon LIF withdrawal and this would be consistent with the high levels of *Nanog* observed in these cells (Figure 3.23b and c).

(a)

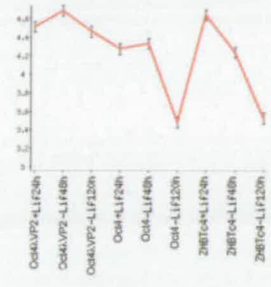
Nanog



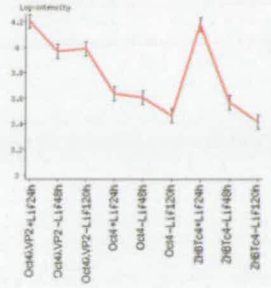
Gdf3



Dppa2

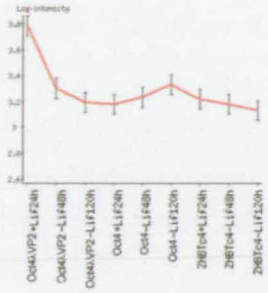


Fbxo15

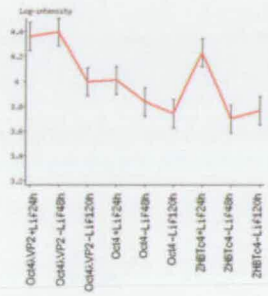


(b)

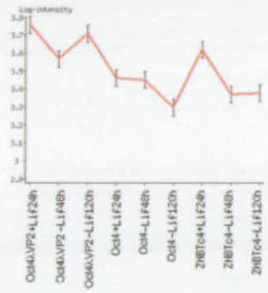
Dppa3



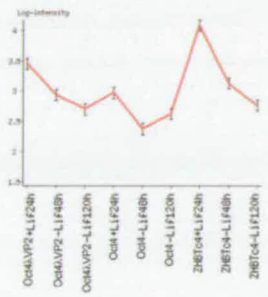
ERas



Sox15

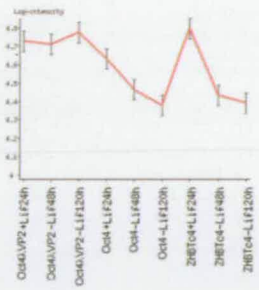


Ecat1

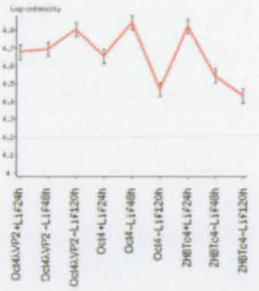


(c)

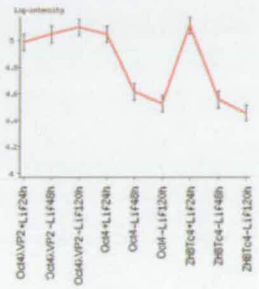
Sox2



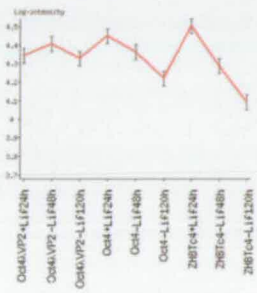
Dnmt3l



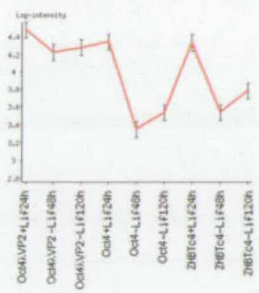
Rex1



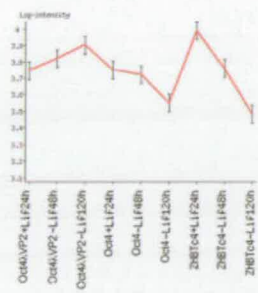
Dppa4



Klf4

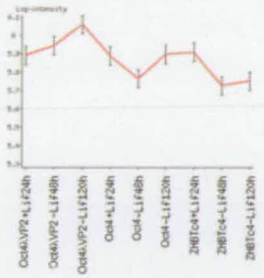


Ecat8

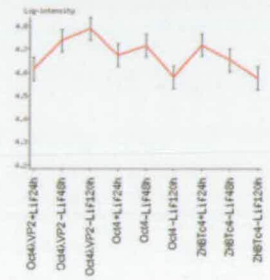


(d)

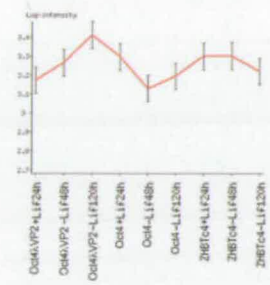
Dppa5



Utf1

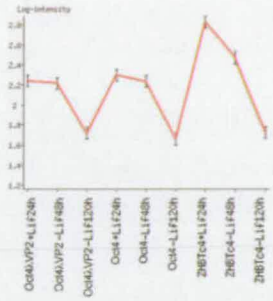


Ctnnb1

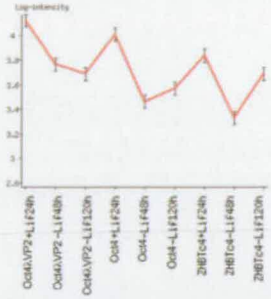


(e)

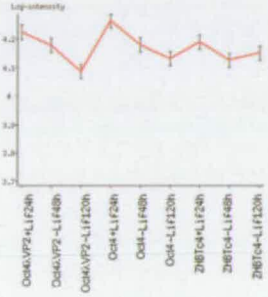
Fthl17



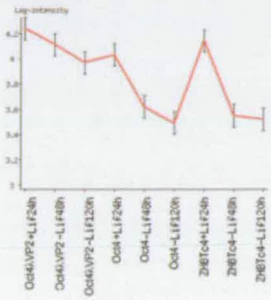
Stat3



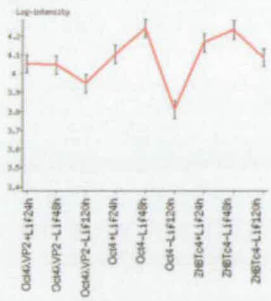
Grb2



Tcl1



c-Myc



(f)

Sall4

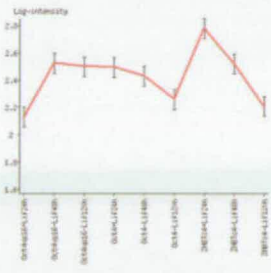


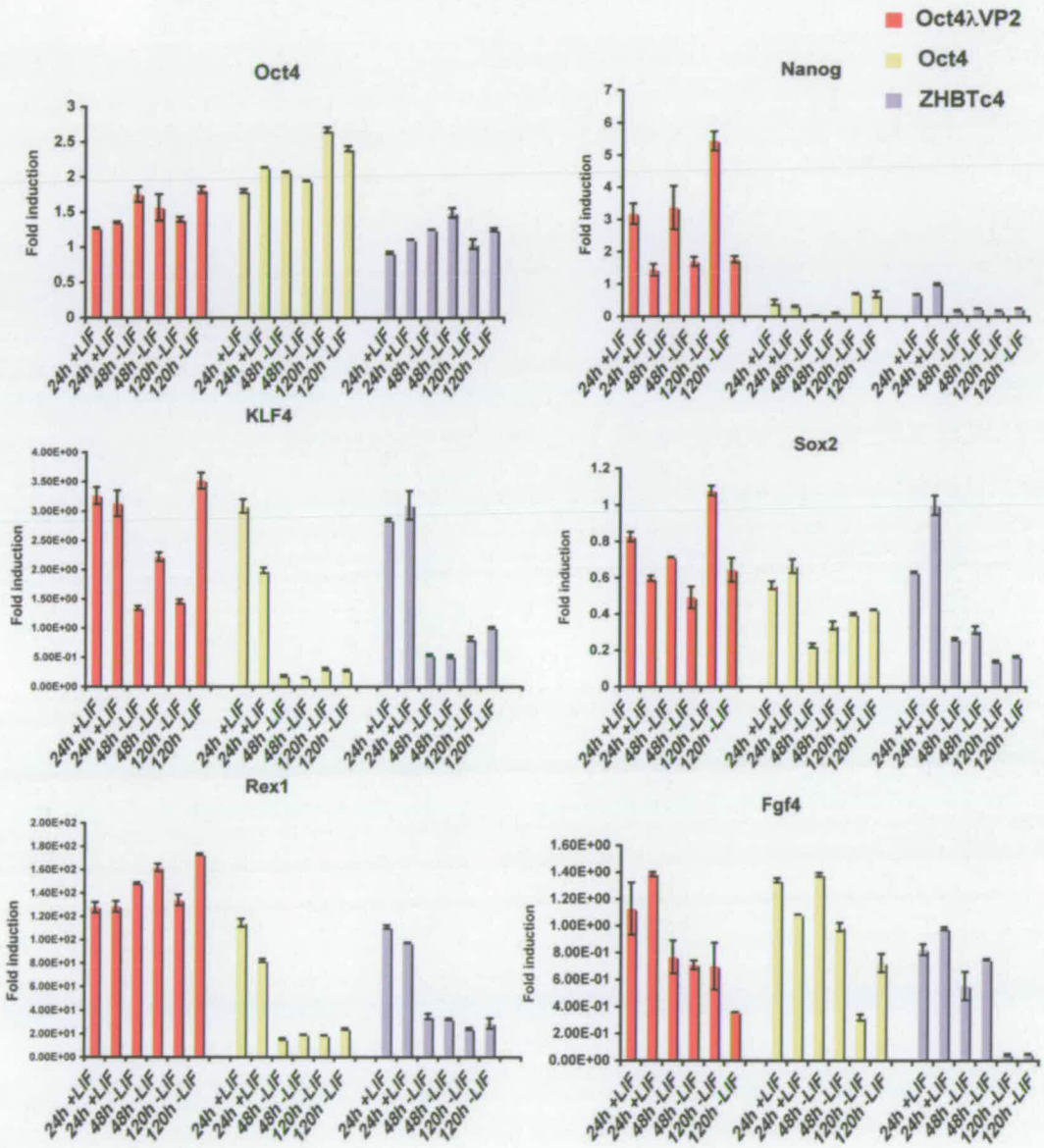
Figure 3.22-Gene expression analysis of ES Cells Associated Transcripts (ECAT) in Oct4λVP2 cell line during LIF withdrawal time course

Microarray plots: the plots are comparing log intensity values for each gene probe among the three cell lines at the three time points of the LIF withdrawal experiment. The error bars represents standard deviation between two biological replicates (two independent clones).

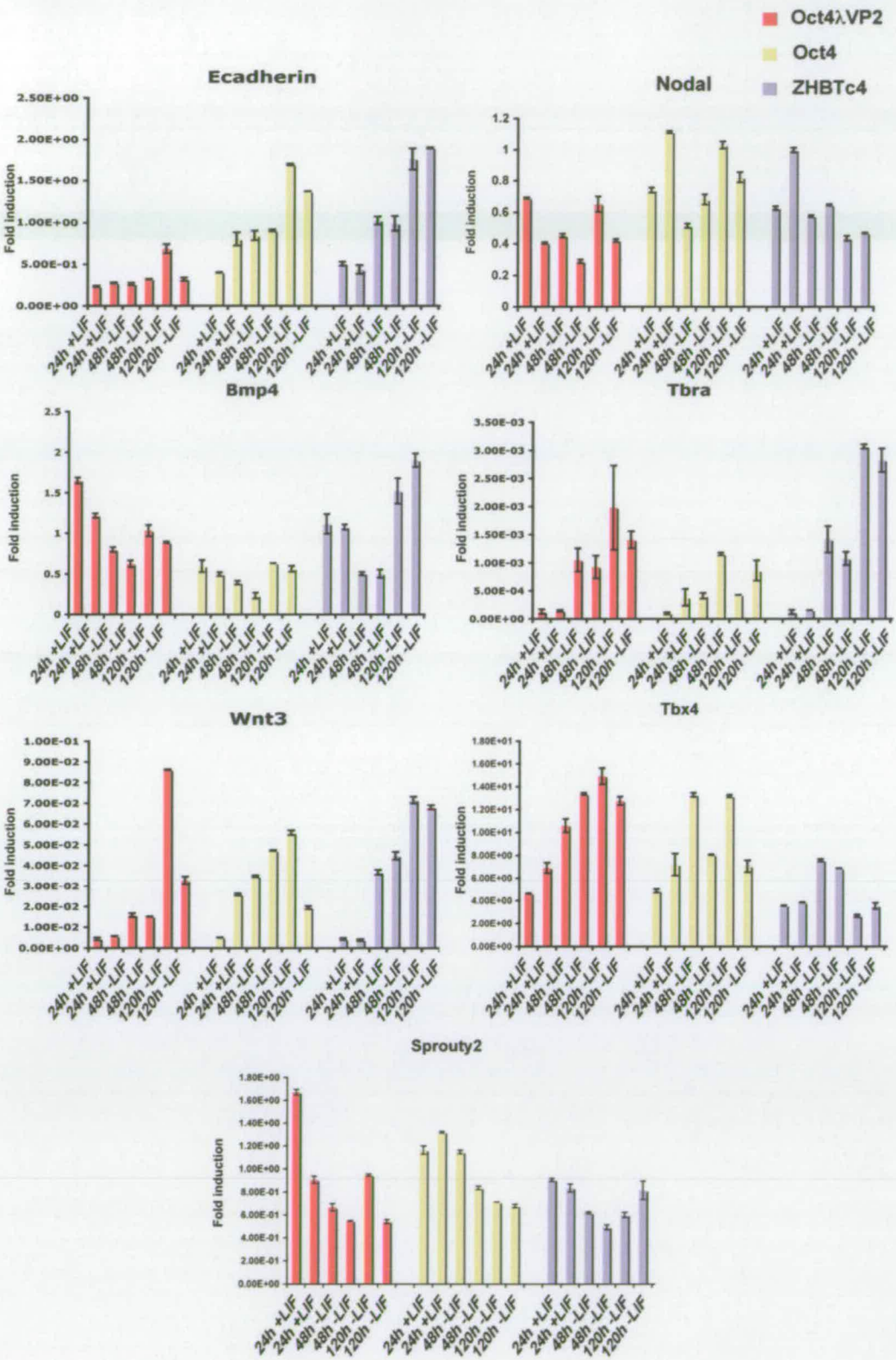
Total RNA was collected during the LIF withdrawal time course as explained previously in the microarray experiment design.

(a) Upregulated in Oct4λVP2 cell line and are maintained after LIF withdrawal.
(b) Upregulated in Oct4λVP2 cell line and are downregulated after LIF withdrawal.
(c) Expressed at similar level in Oct4λVP2 cell line and are maintained after LIF withdrawal.
(d) Expressed at similar level in Oct4λVP2 cell line and are upregulated after LIF withdrawal.
(e) Expressed at similar level in Oct4λVP2 cell line and are downegulated after LIF withdrawal.
(f) Expressed at lower levels in Oct4λVP2 cell line.

(a)


Oct4λVP2
Oct4
ZHBTc4
Oct4
Nanog
KLF4
Sox2
Rex1
Fgf4
Fold induction
24h +LIF
24h +LIF
48h -LIF
48h -LIF
120h -LIF
120h -LIF

(b)


Oct4λVP2
Oct4
ZHBTc4
Ecadherin
Nodal
Bmp4
Tbra
Wnt3
Tbx4
Sprouty2
Fold induction
24h +LIF
24h -LIF
48h -LIF
120h -LIF

(c)

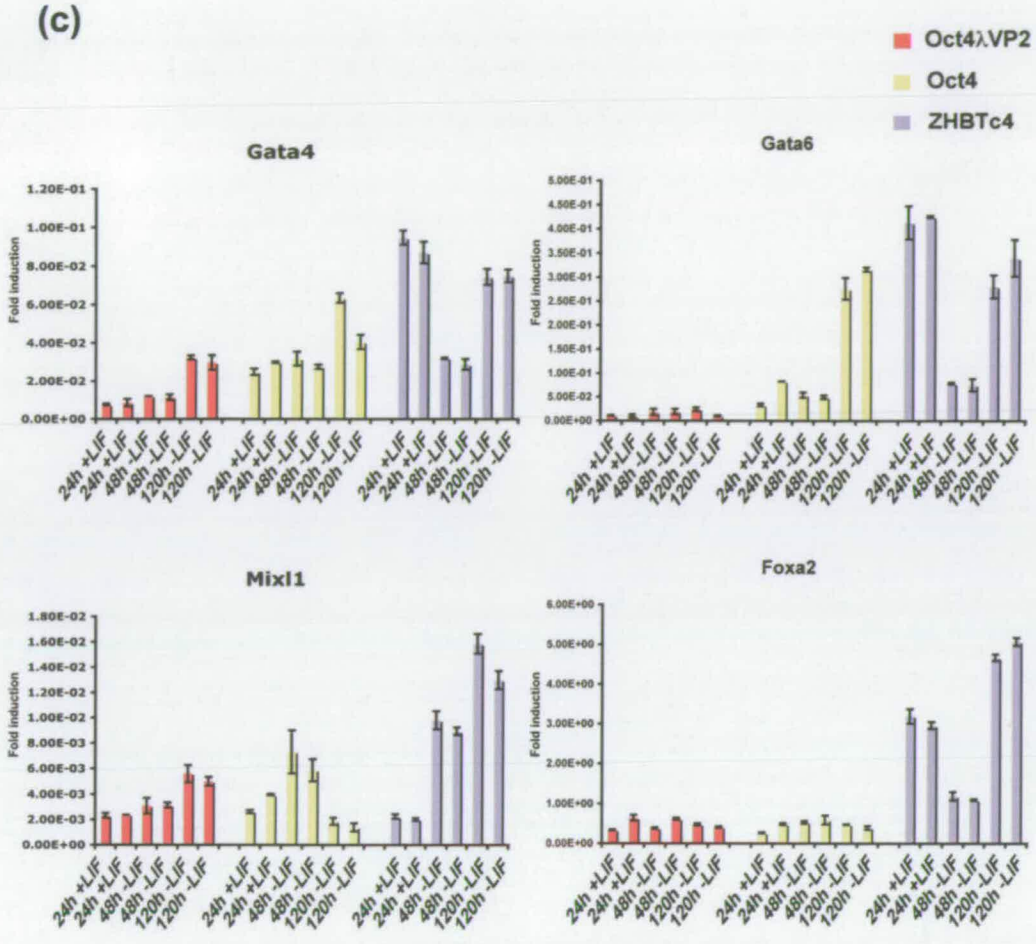


Figure 3.23-Gene expression analysis of Oct4λVP2 cell line during LIF withdrawal time course

Total RNA was collected during the LIF withdrawal time course as explained previously in the microarray experiment design.

(a) qRT-PCR analysis of pluripotency genes expression levels in Oct4λVP2 cell line during LIF withdrawal time course: Oct4, Nanog, KLF4, Sox2, Rex1 and Fgf4 expression levels remain high even after LIF withdrawal.

(b) and **(c)** qRT-PCR analysis of expression levels of mesoderm and endoderm markers in Oct4λVP2 cell line during LIF withdrawal time course: GATA4 and GATA 6 are significantly downregulated .
Expression levels of all analysed genes were normalised to TBP levels for each sample.

3.2.6.5. Correlation between *Nanog* levels and the ability of activator supported cell line to differentiate upon LIF withdrawal

The expression of *Nanog* in Oct4λVP2-supported cell line suggests that Oct4λVP2 regulates *Nanog* levels. We therefore asked what would be the consequences of reducing *Oct4λVP2* levels on *Nanog* expression and on the response of Oct4λVP2-supported cell line to LIF withdrawal. As *Oct4λVP2* expression is selectable based on an IRES puromycin cassette downstream of the *Oct4λVP2* coding sequence, one way to reduce its expression level was to modify the concentration of puromycin used for selection. Thus, we generated a new set of cell lines by selecting for transgene expression at lower puromycin concentrations. The resulting cell lines (Oct4λVP2*) express lower levels of *Oct4λVP2* (Figure 3.24, first panel) than the previously derived cell lines, and, also express correspondingly lower levels of *Nanog* (Figure 3.24). This reduction of both *Oct4λVP2*, and *Nanog* levels in Oct4λVP2* cell lines was confirmed at the protein level using immunostaining (Figure 3.25). While there seems to be modest effects on other pluripotency factors at an RNA level, only *Nanog* shows a consistent greater than two fold change in two independent clonal lines.

When the newly derived Oct4λVP2*-supported cell lines were plated at clonal density and allowed to expand. The formed colonies exhibited an intermediate phenotype between the small, highly undifferentiated phenotype of the original Oct4λVP2-supported lines and the typical ES cell colonies formed by any of the control cell lines. Colonies derived from Oct4λVP2* cells were slightly bigger than those derived from the original Oct4λVP2 cell line, with more differentiated cells in their periphery (Figure 3.26a). Moreover, self-renewal rate of Oct4λVP2* cells was comparable to that of Oct4λVP2, Oct4 or ZHBTc4 cell lines (Figure 3.26b). The Oct4λVP2* cell line also expressed key pluripotency associated genes such as *Klf4*, *Sox2*, *Rex1*, and *Fgf4* at similar levels to Oct4λVP2, Oct4, and ZHBTc4 cell lines (Figure 3.24). Taken together, our findings suggest that the Oct4λVP2* cells have an entirely normal undifferentiated phenotype.

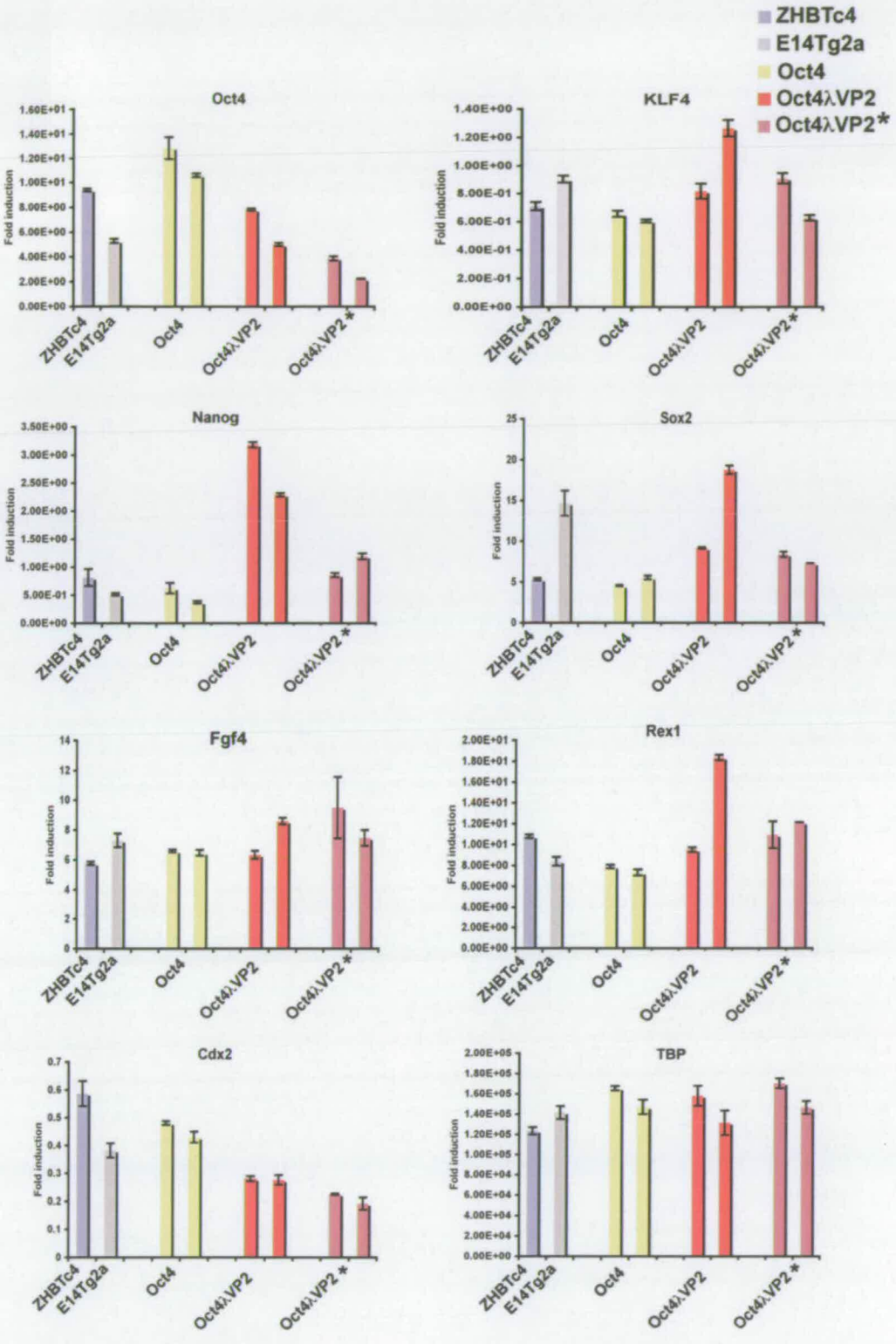


Figure 3.24-Oct4λVP2* cell line express Oct4λVP2 and Nanog at lower levels than Oct4λVP2 cell line under self-renewing conditions

Total RNA collected from the indicated cell lines was used for qRT-PCR analysis of expression levels of pluripotency markers and trophoblast differentiation marker Cdx2.
Expression levels of all analysed genes were normalised to TBP levels for each sample.

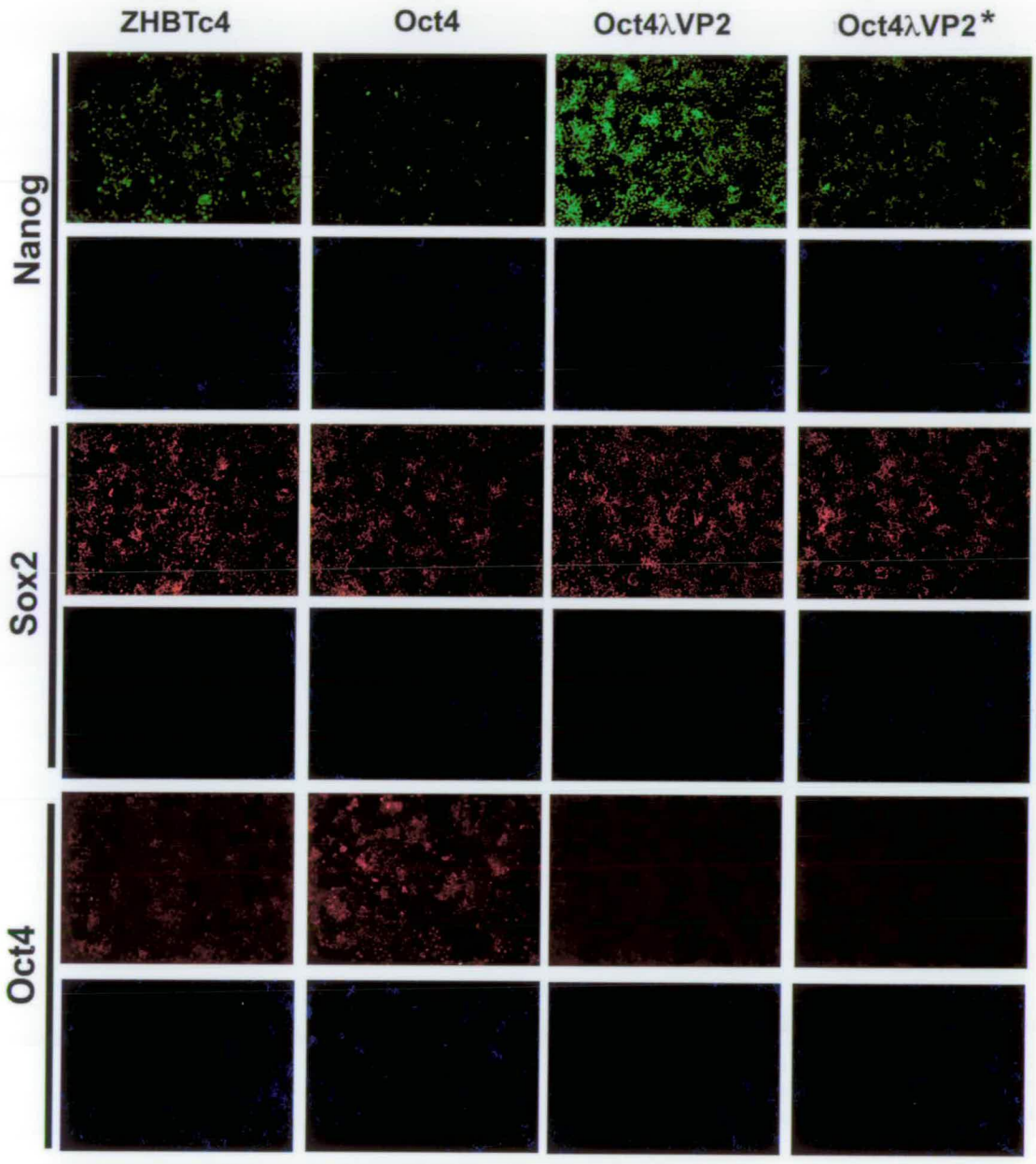


Figure 3.25-Immunocytochemistry of Oct4λVP2* cell line under self-renewing conditions

Three clones from each cell line (except ZHBTc4) cultured under self-renewing conditions were fixed and immunostained for Oct4, Nanog and Sox2 proteins. Dapi nuclear staining (blue field) images are shown for each cell line and each antibody.

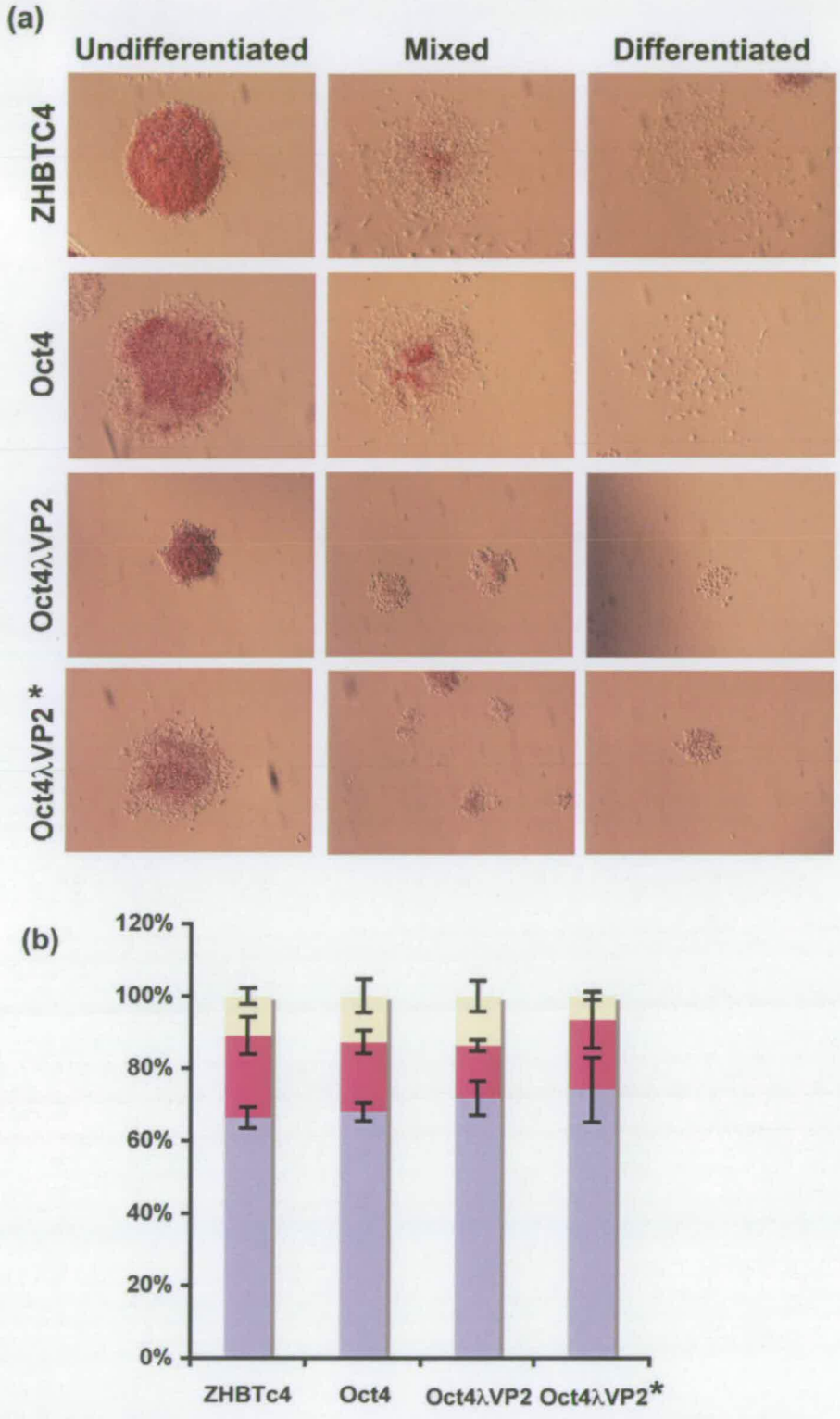


Figure 3.26-Self-renewal assay of newly derived Oct4λVP2* cell line

The indicated cell lines were plated at clonal density (60 cells/cm2), and cultured for 6 days in self-renewing condition then stained for alkaline phosphatase (AP) activity (red).

(a) Examples of colony morphology present in each cell line. Colonies were classified into three categories: Uniformly AP positive undifferentiated colonies, mixed colonies containing AP positive and negative cells, and AP negative differentiated colonies.

(b) Percentage of previous colony categories in each cell line. Data represents average values from two independent clones from each cell line.

To determine whether these activator supported cell lines with reduced Nanog were now able to differentiate, we plated them clonally and asked whether they maintained undifferentiated morphology when challenged by either Oct4 overexpression or LIF withdrawal. Figure 3.27 shows that Oct4λVP2* cells, unlike Oct4λVP2 cells, differentiated in the absence of LIF. Interestingly, these cell lines remained at least partially resistant to the effect of Oct4 overexpression.

Taken together these observations suggest that the capacity of Oct4 activator supported cells to differentiate is directly proportional to the level of the activator transgene, and based on our limited analysis, the single clearest response in undifferentiated ES cells to this transgene appears to be Nanog expression. However while Oct4λVP2* cells express similar levels of *Nanog* as the wild type control cell lines (Figure 3.24), they are still relatively immune to Oct4 overexpression, suggesting that either Oct4 overexpression induced differentiation is very sensitive to *Nanog* levels or that some aspects of the Oct4λVP2 phenotype are *Nanog* independent.

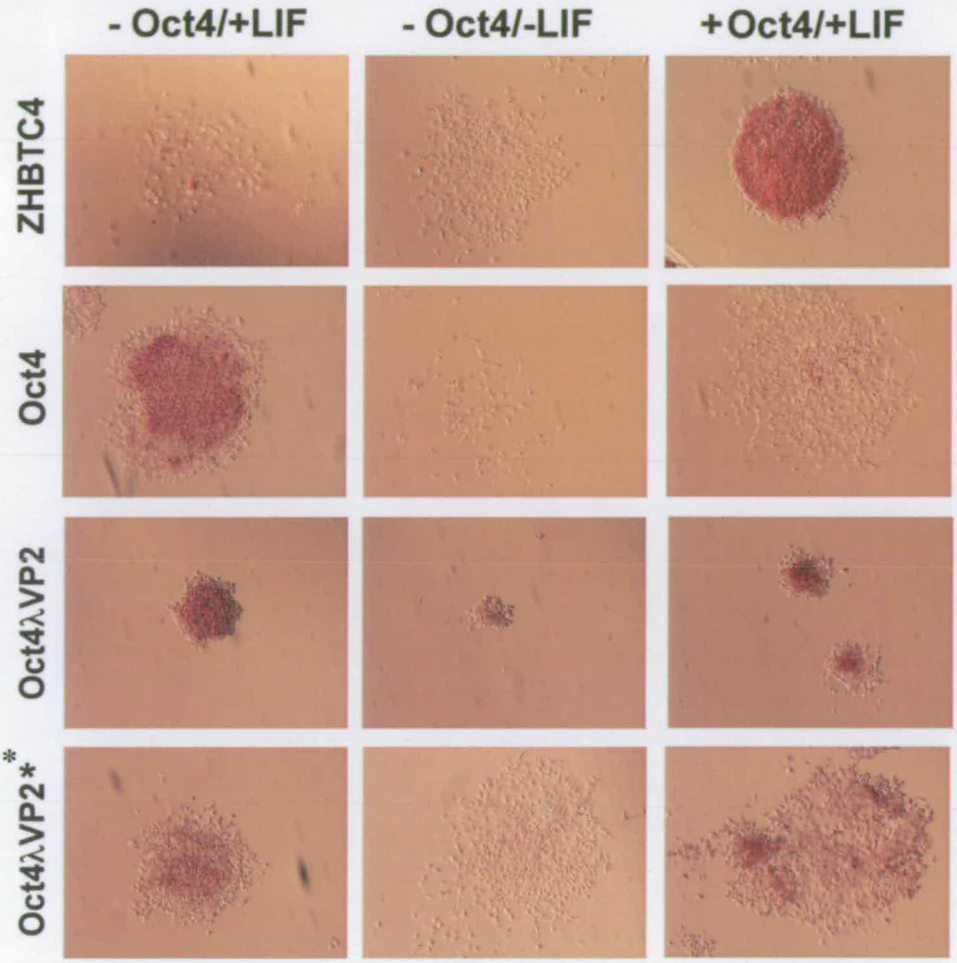


Figure 3.27-Oct4λVP2* cell line is able to differentiate upon LIF withdrawal or Oct4 overexpression

Clonal density differentiation of Oct4λVP2 and Oct4λVP2* cell lines: cells were plated at clonal density in the absence of LIF or Oct4 overexpression condition for 5 days as well as self-renewing conditions as a control. Three clones for each cell line except ZHBTc4 were used and stained for AP activity (red). Photos of representative colonies are shown.
Oct4 overexpression condition is achieved by removal of tetracycline from the medium leading to the re-expression of the Tc responsive Oct4 transgene.

-Oct4/+LIF represents self-renewing condition.
-Oct4/-LIF represents LIF withdrawal condition.
+Oct4/+LIF represents Oct4 overexpression condition.

3.2.7. Identification of potential novel pluripotency genes

We cannot be certain that only *Nanog* upregulation is responsible for the highly undifferentiated phenotype of Oct4λVP2 cell line as we also see similar expression dynamics in other ECAT genes such as *Gdf3* gene. *Gdf3* is linked to pluripotency and co-regulated within the *Nanog* locus by Oct4 (Levasseur et al., 2008). Other key unkown pluripotency genes that are also upregulated and affected by *Oct4λVP2* levels could be contributing to the phenotype of the cells. To limit our analysis to those genes directly downstream of Oct4, we overlapped our dataset with the available ChIP on chip data for Oct4 (Kim et al., 2008; Loh et al., 2006; Sharov et al., 2008) (Figure 3.28).

Of the 980 genes upregulated in Oct4λVP2 supported cells, 147 of them had been identified as potential targets based on ChIP-Seq and ChIP on chip data.
The lists of these 147 genes and their gene ontology analysis are attached as appendices (Table 3.5, and 6). These genes were mainly involved in cell fate commitment, growth, proliferation, differentiation, kinase activities, and developmental processes.

Six of these 147 genes; acid phosphatase6 (*Acp6*), F-box protein 15 (*Fbxo15*), Nanog homeobox (*Nanog*), patched homolog1 (*Ptch1*), transcription factor CP2-like 1 (*Tcfcp2l1*), and ubiquitin-like, containing PHD and RING finger domains 2 (*Uhrf2*) were common to all the previous studies (Table 3.5).
More analysis is currently underway to further refine this list of genes, and investigate their functional relevance.

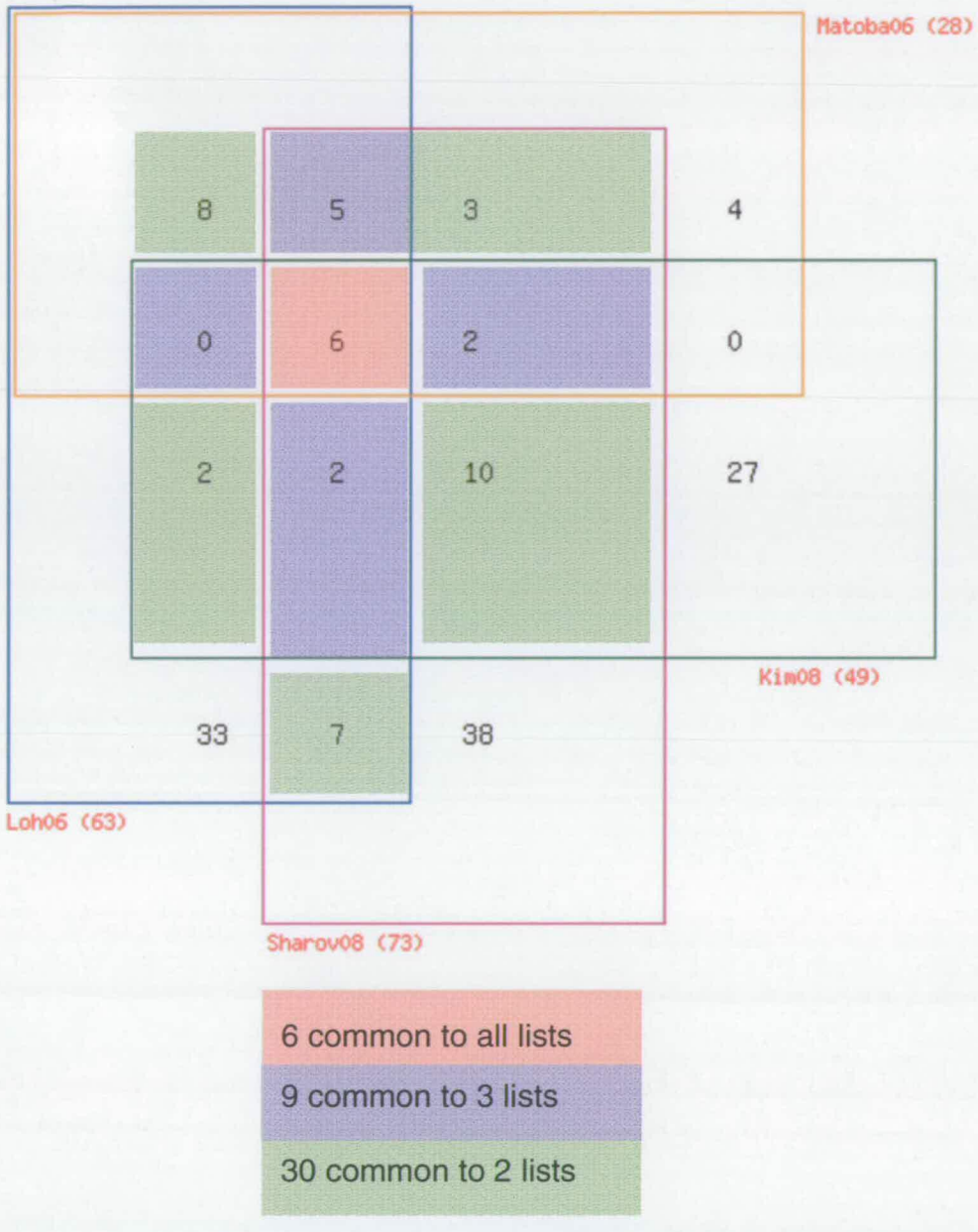


Figure 3.28-Venn diagram showing that 147 genes upregulated in Oct4λVP2 cell line were identified as possible Oct4 targets

List of 980 upregulated genes in Oct4λVP2 cell line under self-renewing conditions was compared to four published lists of predicted Oct4 target genes. The published lists are from studies by Matoba et al 2006, Loh et al 2006. Kim et al 2008, Sharov et al 2008.
Pink colour represents genes tat are common to all the lists
Purple colour represents genes that are common to three lists
Green colour represents genes that are common to two lists.

3.2.8. Genes upregulated in Oct4λVP2 include genes that are either targets for Oct4 only, Nanog only or both

Comparison of the list of 980 upregulated genes in Oct4λVP2 supported cell lines to published lists of predicted Oct4 or Nanog targets from ChIP/chip studies and functional studies allows us to further categorize the upregulated genes into Oct4 only targets, Nanog only targets, and targets of both (Kim et al., 2008; Loh et al., 2006; Matoba et al., 2006; Sharov et al., 2008). We can also superimpose functional constraints on these categories based on the response of these genes to Oct4 or *Nanog* knockdown. When this analysis was performed, we found that a significant number of genes that are upregulated in the Oct4λVP2 supported cell lines are Nanog repressed targets (Table 3.7 appendix). This is particularly interesting as these cells express high levels of *Nanog*. Moreover, a number of these genes are developmental genes and targets of both Oct4 and Nanog. The repressed developmental genes are believed to be regulated by a constellation of factors that include Oct4, Nanog, and Sox2 (Chen et al., 2008b; Kim et al., 2008). The fact that a significant number of these supposedly repressed genes are upregulated in Oct4λVP2 cells supports the notion that Oct4λVP2 acts as an activator despite its presence in this allegedly repressive complex and can override Nanog mediated repression.

3.3. Discussion

In this chapter, I have shown that the major mechanism by which PouV proteins block differentiation is by positively maintaining a network that regulates pluripotency. Aspects of this network appear conserved and able to regulate gastrulation stage lineage specification and ES cell self-renewal. While PouV proteins acting as repressors induce differentiation, the activator forms are able to maintain cells in uncommitted state and block differentiation in both murine ES cells and embryonic progenitor cells in Xenopus. This block of differentiation by PouV activator fusions is regulated indirectly through the upregulation of key factors including *Nanog*.

The notion that direct gene repression by PouV proteins is essential for ES cell self-renewal is at odds with our findings. However, we considered the possibility that the fusion of PouV proteins to either VP16 or EnR domains did not effectively transform them exclusively to activators or repressors, respectively. The use of this approach was pioneered in lower vertebrates. A number of studies have deleted endogenous regulatory (activation or repression) domains from the protein and replaced them with peptide motifs that have the opposite function (Croston et al., 1992; Jaynes and O'Farrell, 1991). However, these studies were limited because the removal of the endogenous domains may have affected the cooperative binding of co-factors and neutralised the regulatory function of the proteins, resulting in alteration of their DNA binding specificity. However, in this study we have not removed any endogenous PouV protein sequence, and therefore a loss in DNA binding specificity is unlikely. We have used this approach previously and it appears to be the most effective way to preserve DNA binding specificity and convert regulatory function (Brickman et al., 2001; Brickman et al., 2000). The use of these fusions with Hex led to the identification of a short list of targets in ES cells and these were later validated in development, indicating that there was no alterations in specificity (Zamparini, Watts et al. 2006). Furthermore, there are several technical reasons pointing towards the fact that VP16 and EnR domains will promote strong trans-activation or trans-repression without affecting DNA binding specificity of the fused transcription

factor. The inclusion of the flexible λ linker between the VP16 or EnR domain and the wild type protein allows it to fold correctly and minimise changes to its tertiary structure and therefore reduces the likelihood of the activator or repressor domain to interfere with its binding to DNA. Although it is possible that the addition of VP16 or EnR heterologous domains to the PouV proteins have a gain of function effect on DNA binding, fusing either of these domains to a DNA binding deficient Oct4, generated by a single amino acid change in the POU DNA binding domain, produced proteins that were unable to activate or repress transcription from reporters genes harbouring PouV responsive elements. Suggesting that the only DNA binding ability of these fusion proteins is restricted to Oct4 POU domain, and therefore it is unlikely that VP16 or EnR domains affect Oct4 DNA binding specificity. However, to fully exclude the possibility of gain or loss of DNA binding caused by fusing VP16 or EnR domains to the wild type protein, a genome wide ChIP analysis (ChIP/Chip or ChIP-seq) of the wild type and the fusion proteins is necessary. Oct4λVP2/Oct4λEnR ChIP/Chip will determine whether they bind the same sites as wild type Oct4. Moreover, combining this data with expression profiles will delineate whether they regulate the same genes as Oct4, and therefore allow investigation of their ability to bind co-factors associated with wild type Oct4.

While we are reasonably confident that the use of our modular domain system with the built-in flexibility of the l linker will convert full length transcription factors from activators to repressors and vice versa, we extensively tested the PouV fusion proteins used in this study. We concluded that the activator fusion proteins had no residual repression activity at three levels; 1. Reporter assays harbouring Oct4 responsive elements. None of our activator fusions was able to repress the basal level transcription from the *tk* promoter, whereas engrailed fusions did. 2. The Oct4λVP2 fusion protein was unable to recapitulate the direct inhibition of the β-Subunit of Human Chorionic Gonadotropin (*hCG*-β) promoter by wild type Oct4 in Jar human choriocarcinoma cells. 3. In ES cells stably expressing Oct4λVP2, promoters normally co-occupied by Oct4 and Nanog and repressed by Nanog are still upregulated. Therefore, even in the context of the global repression complex proposed by Liang et al (2008), Oct4λVP2 acts as an activator and overrides Nanog

repression despite the latter being expressed at high levels in this cell line (Liang et al., 2008).

While we can never be sure that we have effectively converted Oct4 from a repressor to an activator at every promoter it normally regulates, we conclude that Oct4λVP2 is an effective activator of gene transcription. In fact, the ability of Oct4λVP2 to recapitulate the function of wild type Oct4 and to regulate most of the genes in the same manner support the idea of Oct4 being mainly an activator and rarely a repressor of gene expression. Therefore, Oct4 suppresses lineage commitment indirectly. This is in agreement with what has been reported by Sharov et al (2008). The majority of Oct4 tentative target genes identified in the study are potent repressors of transcription that were downregulated after Oct4 suppression (Sharov, Masui et al. 2008).

The idea that direct repression of genes important for lineage commitment by Oct4 is essential for maintaining undifferentiated ES cells is based on the following observations: 1) the upregulation of differentiation specific genes following Oct4 knockdown (Hay et al., 2004); 2) these genes that are upregulated following *Oct4* knockdown possess octamer binding sites in their regulatory regions or can ChIP Oct4 (Chen, Xu et al. 2008; Kim, Chu et al. 2008).

The fact that Oct4 acts mainly as an activator and suppresses lineage commitment indirectly questions the functional significance of Oct4 repressed targets in undifferentiated cells. These repressed or silent targets could be non-functional false positive targets that were a result of random binding of Oct4 to Octamer motifs present in most class II promoters. Alternatively, they could be genuine Oct4 targets, whose repression is not necessary for maintaining uncommitted cells and blocking differentiation. These targets may not be active in ES cells, but could be potential targets of Oct4 later in differentiation. The abundance of neural genes in all the Oct4 target data sets supports the notion that the set of functional Oct4 targets may change as development progresses.

Strikingly when we assess the complete set of genes that respond negatively to Oct4 in ES cells, the majority of them are downregulated by expression of Oct4λVP2, indicating that some aspect of the Oct4 dependent network keeps these genes off.

Many observations support the possibility of Bmp4 being the effector downstream of PouV proteins to block differentiation. *Bmp4* expression is dependent on PouV expression in Xenopus and is marginally upregulated in response to Oct4λVP2 in ES cells. Moreover, its overexpression in *PouV* depleted Xenopus embryos restored the normal gene expression pattern of *Xvent2* and *Goosecoid.* Another circumstantial observation implicating Bmp4 comes from the fact that its downstream target *Xvent2* (also known as Xom), an immediate early target of Bmp signalling, is also predictably PouV dependent as it has been shown to be a target of Xlpou25 (Cao, Knochel et al. 2004). Although this data indicates that Bmp4 might be the effector of PouV proteins to suppress differentiation, more analysis is required to confirm this. In addition to its role in Xenopus development, Bmp4 is also associated with maintaining pluripotency and suppressing differentiation in mouse ES cells. It was identified as the component of serum required to maintain ES cells in the presence of LIF and absence of feeders (Ying, Nichols et al. 2003).

The fact that *Nanog* was the only gene amongst the core pluripotency genes that was highly upregulated in Oct4λVP2 cell line could be explained by a random ability of the cells to tolerate only high levels of Nanog. However, this seems an unlikely explanation mainly because expression levels of Oct4λVP2 are positively correlated only with *Nanog* levels amongst all the key pluripotency factors. This correlation suggests the existence of an Oct4-Nanog pathway, and indicates that in mouse ES cells Oct4 might be exerting its function mainly through Nanog and factors downstream.

As Nanog may be acting as a master transcriptional organizer during reprogramming to instill pluripotency (Silva, Chambers et al. 2006; Silva, Barrandon et al. 2008; Silva, Nichols et al. 2009), its ectopic expression in response to Oct4λVP2 maybe locking cells in a pluripotent cycle.

CHAPTER 4

Mapping functional domains of Oct4 homologues responsible for ES self-renewal

4.1. Introduction

Oct4 is a transcription factor that plays a major role in regulating pluripotency and self-renewal in both embryos and embryonic stem cells. It is also one of the four factors able to induce the reprogramming of somatic cells to pluripotency and appears to be the only one indispensable in most reprogramming experiments (Do and Scholer, 2009).

Oct4 belongs to the POU family of transcription factors. This protein family can be broadly broken down into seven classes based on sequence homology of their POU domains and their linker region (Wegner, Drolet et al. 1993; Spaniol, Bornmann et al. 1996). (For more details, please refer to the section of Oct4 structure in the general introduction).

Interestingly, within the PouV sub-class, proteins differ in their ability to substitute for Oct4 function in the maintenance of ES cell self-renewal. While these proteins differ in their ability to support self-renewal, they appear to have equivalent activity in transient reporters assays (Morrison and Brickman, 2006). These findings suggest that the capacity of these proteins to recognize DNA and activate transcription is the same, and the difference in their activities might depend on the presence or the absence of specific protein-protein interaction motifs required for Oct4 to recognize complex promoters in vivo.

Oct4 has been structurally divided into three regions; a bipartite DNA-binding domain consisting of the POU-specific and the POU-homeo domains joined by a 17 amino acid linker, and both an amino-terminal (N-TD) and a carboxyl-terminal (C-TD) transactivation domains (Pan et al., 2002). The N-TD is classified as proline rich whereas C-TD is a serine/threonine rich with high proline content (Okamoto, Okazawa et al. 1990; Rosner, Vigano et al. 1990; Scholer 1991). Several studies revealed functional difference between the amino and carboxyl terminal transactivation domains. Some reported detection of transactivation ability only in the N-terminus (Imagawa et al., 1991), whereas others reported that both N-TD and

C-TD are active (Vigano and Staudt 1996; Brehm, Ohbo et al. 1997). Furthermore, the activity of these transactivation domains is affected differently by protein-protein interaction (Ambrosetti, Scholer et al. 2000), and is regulated in a cell type dependent manner (Brehm, Ohbo et al. 1997). The most relevant study addressing the identification of Oct4 specific domains responsible for ES cell self-renewal was done by Niwa and his colleagues (2002). They found that both the N-TD and the C-TD share redundant functions in ES self-renewal, and concluded that the POU domain of Oct4 is uniquely required to maintain the undifferentiated phenotype of ES cells (Niwa et al., 2002).

The relevance of these different domains to the specific ability of certain PouV proteins to rescue *Oct4* null ES cells is not clear. Therefore, we exploited the findings of a previous study in the lab assessing the ability of different classV POU proteins from different species to substitute for Oct4 function in mouse ES cells (Morrison and Brickman, 2006) in order to confirm and further refine the domains that are necessary for ES cell self-renewal. In this chapter, I show that PouV proteins able to rescue ES cell self-renewal appear to have two separable activities, a quantitative capacity to support clonal growth that is regulated through the C-TD and the support of ES cell morphology through the POU domain.

4.2. Results

4.2.1. PouV protein sequence analysis and the construction of chimeric proteins

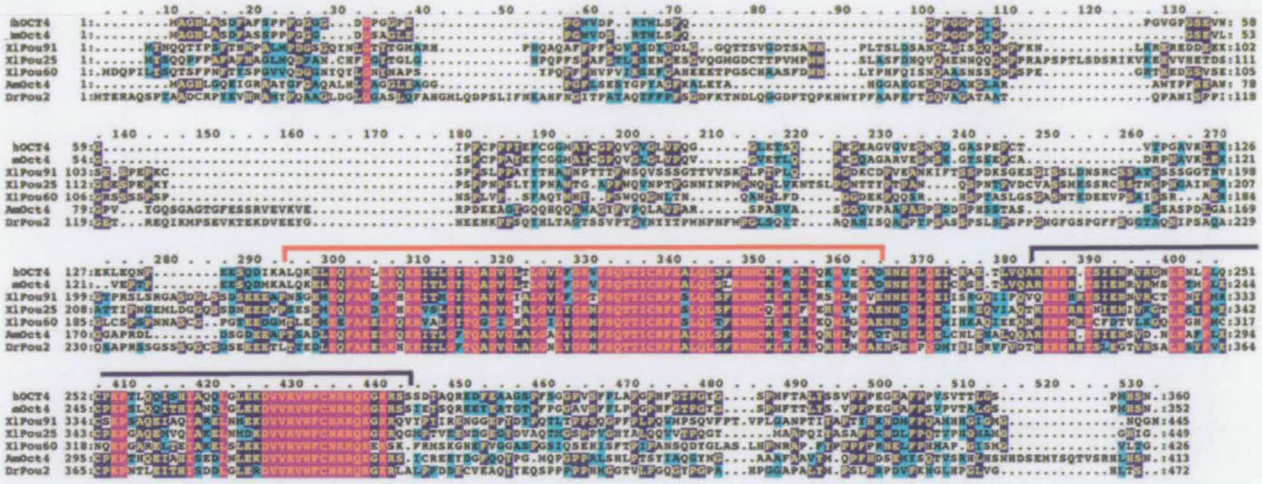
As previously stated, recent studies in the lab, revealed that PouV proteins vary greatly in their ability to maintain ES cell self-renewal in spite of their sequence similarity. While the three Xenopus PouV proteins (Xlpou91, Xlpou60, and Xlpou25) are highly related to each other (based on the percentage of sequence identity of their various domains), they differ in their capacity to substitute for mouse Oct4 activity in the support of murine ES cells (Morrison and Brickman, 2006). Moreover, the study also showed that the zebrafish PouV protein, DrPou2 has no activity at all in this assay.

An experimental strategy to identify functional domains responsible for maintaining the undifferentiated phenotype of ES cells was designed exploiting the sequence similarity and the functional differences of PouV proteins. Based on homology in the POU domain (both POU specific and POU homeodomain) (Figure 4.1a), we have constructed a series of chimeric proteins by replacing domains from the two proteins that have only low levels of Oct4-like activity, Xlpou25 and DrPou2, with those from the Xlpou91 protein with the Oct4-like activity (Figure 4.1b).

While the original DrPou2 had no ability to rescue *Oct4* deficient ES cells, a second zebrafish DrPou2 EST sequence was identified during the course of this study. This variant behaved in a similar manner to Xlpou25 and possessed a low, but detectable ability to rescue ES cell self-renewal.
The difference between the two variants of DrPou2 (DrPou2(T) and DrPou2(A)) is the substitution of one amino acid residue in the POU domain. This will be discussed in more details in later sections of the results.

The swaps (based on Figure 4.1b) essentially divided the proteins into three regions, the amino (N-TD), carboxyl (C-TD) and POU DNA binding domain (POU). The chimeric proteins generated by N-TD, C-TD, or POU domain swapping between Xlpou91 and Xlpou25 or Xlpou91 and DrPou2 are schematically represented in a Figure 4.2, and 4.4. The nomenclature used to refer to these chimeric proteins is based on identifying the origin of each domain in the protein, e.g. N2P91C91, has the N-TD from DrPou2 fused to the POU domain and C-TD from Xlpou91, and N91P25C91 has the N-TD from Xlpou91, the POU domain from Xlpou25, and the C-TD from Xlpou91. The designation for each chimera is listed in the figure next to the schematics (Figures 4.2a, and 4.4a).

(a)



(b)

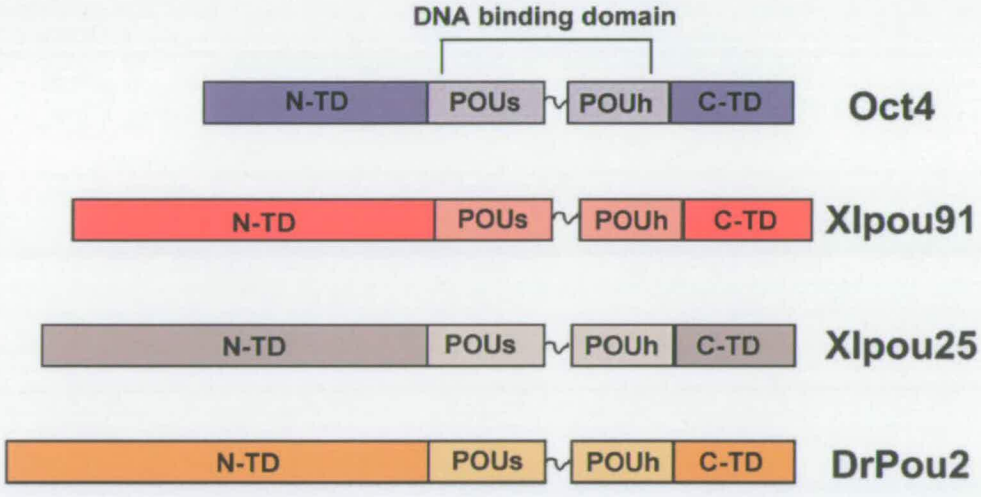


Figure 4.1-PouV proteins: sequence alignments and schematic representation

(a) Alignment of human (hOCT4), mouse (mOct4), Xenopus (Xlpou91, Xlpou25, Xlpou60), axolotl (AmOct4) and zebrafish (DrPou2) PouV proteins. (Alignment from Morrison and Brickman., 2006). POU specific (POUs) and POU homeodomain (POUh) domains are boxed in red and blue respectively. Pink indicates identical residues, light blue indicates similar residues, and dark blue indicates conservation within a subset of residues.
(b) Schematic representations of PouV proteins, and POU domain boundaries.

4.2.2. Measuring the extent of ES cells self-renewal rescue by chimeric PouV proteins

The activity of PouV chimeric proteins was tested at two levels. Firstly, their ability to recognize and activate transcription from Oct4 responsive reporter genes in transient transfection was tested in order to reveal any generic changes in DNA binding and transcriptional activation. Secondly, they were tested in an ES cell complementation assay in order to assess their ability to substitute for Oct4 function in ES cells (Niwa et al., 2002). As outlined in the previous chapter, the complementation assay involves the introduction of PouV transgenes into the ZHBTc4 cell line, an *Oct4* null ES cell line that is supported by a randomly integrated tetracycline regulatable *Oct4* transgene that can be repressed by the addition of tetracycline (Tc) to the media. This system is based on the fact that undifferentiated colonies can form in the presence of Tc only when the exogenous PouV transcription can substitute for Oct4 activity.

The ES cell self-renewal rescue ability of the various chimeric proteins is measured in two ways; short-term and long-term. The short-term self-renewal rescue ability of the chimeric proteins is determined by the number of rescued ES-like alkaline phosphatase (AP) positive colonies generated in the presence of Tetracycline. Whereas, the long-term self-renewal rescue ability is evaluated through the generation of self-renewing, undifferentiated clonal cell lines from the rescued ES-like colonies.

4.2.3. Transcriptional activity of Xlpou91 and Xlpou25/DrPou2 chimeric proteins

The difference in the ability of PouV protein to support self-renewal is not caused by a difference in their ability to activate transcription (Morrison and Brickman, 2006). Therefore, to ensure that differences seen in the ability of the different PouV chimeric protein to rescue ES cell self-renewal is not due to a defect in their

transcriptional regulatory activity, we tested their ability to stimulate transcription from reporter genes harbouring PouV responsive elements.

All the chimeric protein constructs were transiently cotransfected with different reporter genes into ZHBTc4 cells in the presence of Tc (absence of *Oct4* transgene expression). All the reporter genes assayed employed a minimal tk promoter driving luciferase downstream of Oct4 responsive elements either from the *Fgf4* enhancer, 6 reiterated copies of the octamer binding motif (Niwa et al., 2002).

Figure 4.2b1 shows that all Xlpou91-Xlpou25 chimeric proteins activate transcription from the luciferase reporter containing six copies of the octamer binding motif in varying degrees. However, all are at least as active as Oct4 and none appeared to be more active than Xlpou91, which indicates that all these PouV chimeras have transactivation abilities that lie within a transcriptional spectrum that allow the support of ES cell self-renewal.

Moreover, all Xlpou91-Xlpou25 chimeric proteins stimulate similar levels of transcription from the luciferase reporter harbouring elements from the *Fgf4* enhancer (Figure 4.2b2), which further indicates that any differences in the phenotype of rescued colonies by these chimeric proteins is not due to their overall activity levels.

The specificity of the transcriptional activity of these chimeric proteins was confirmed as none of them activated the reporter gene activity in the absence of an Oct4 response element (Figure 4.2b3).

In the same way, Xlpou91-DrPou2(T) chimeric proteins (Figure 4.3) were also tested for their ability to activate transcription. While, these proteins are able to activate transcription at similar levels to Oct4 and Xlpou91 from the luciferase reporter containing six copies of the octamer binding motif (Figure 4.3b1), they were slightly less active on the *Fgf4* enhancer (Figure 4.3b2). However, this slight difference does not appear to correlate with rescue ability of the proteins (see below).

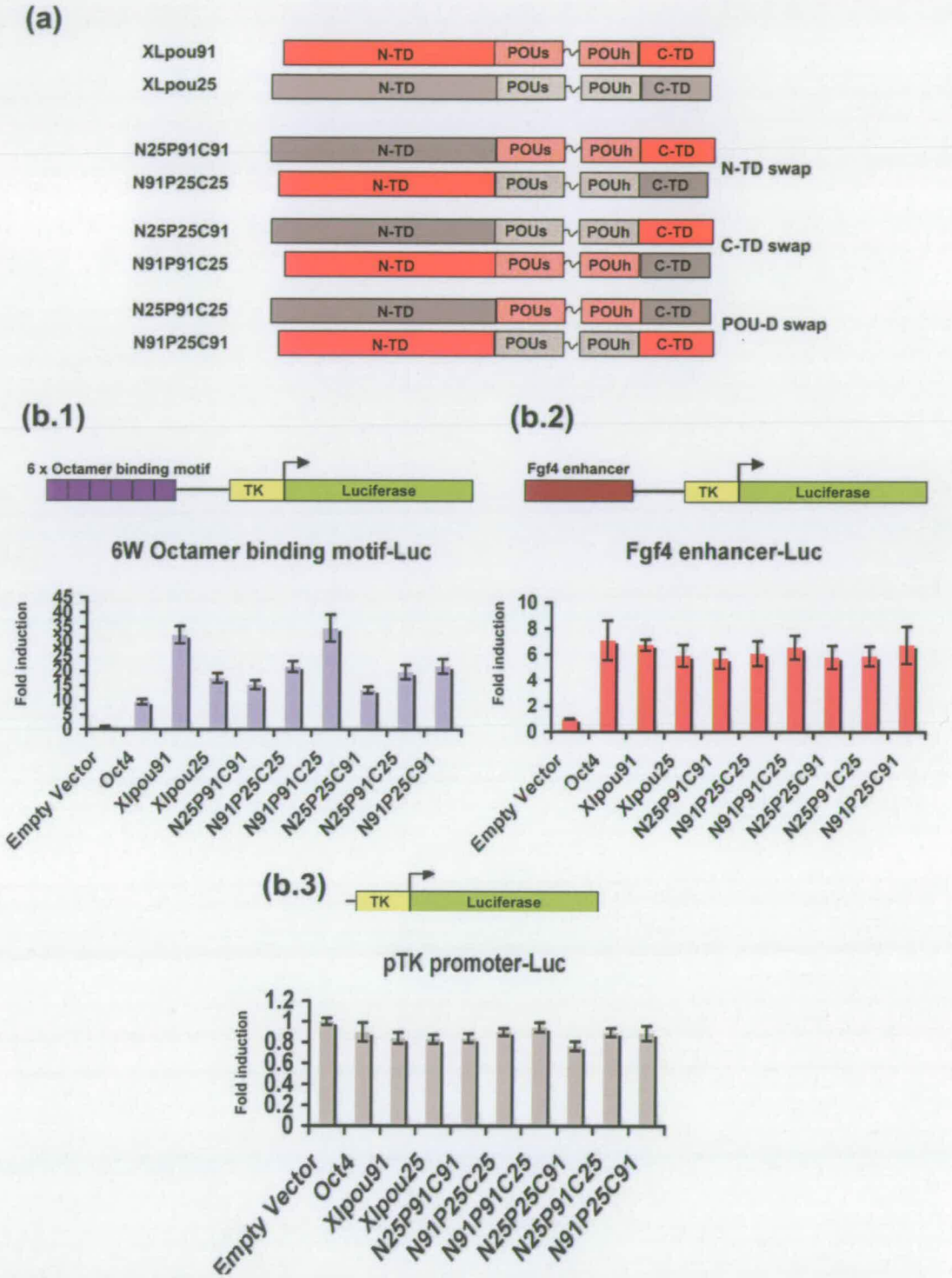


Figure 4.2-Cloning and transcriptional activity of Xlpou91-Xlpou25 Chimeric proteins

(a) Schematic representation of Xlpou91-Xlpou25 chimeric proteins: chimeric proteins were generated by swapping either the amino acid terminal (N-TD), the carboxyl terminal (C-TD) or the POU domain of Xlpou91 with Xlpou25 domains. **(b)** Xlpou91-Xlpou25 chimeric proteins are able to activate transcription in a specific manner from two luciferase reporters: one containing multiple copies of the octamer binding motif **(b.1)** and the other one containing the Fgf4 enhancer **(b.2).** A luciferase reporter containing the thymidine kinase (TK) promoter was used to test the specificity of the chimeric proteins transcriptional activity **(b.3)**.
The indicated cDNAs were cotransfected with luciferase reporters. Fold induction represents the increase in transcription compared with the empty vector control. Data represent the mean value of two independent experiments.

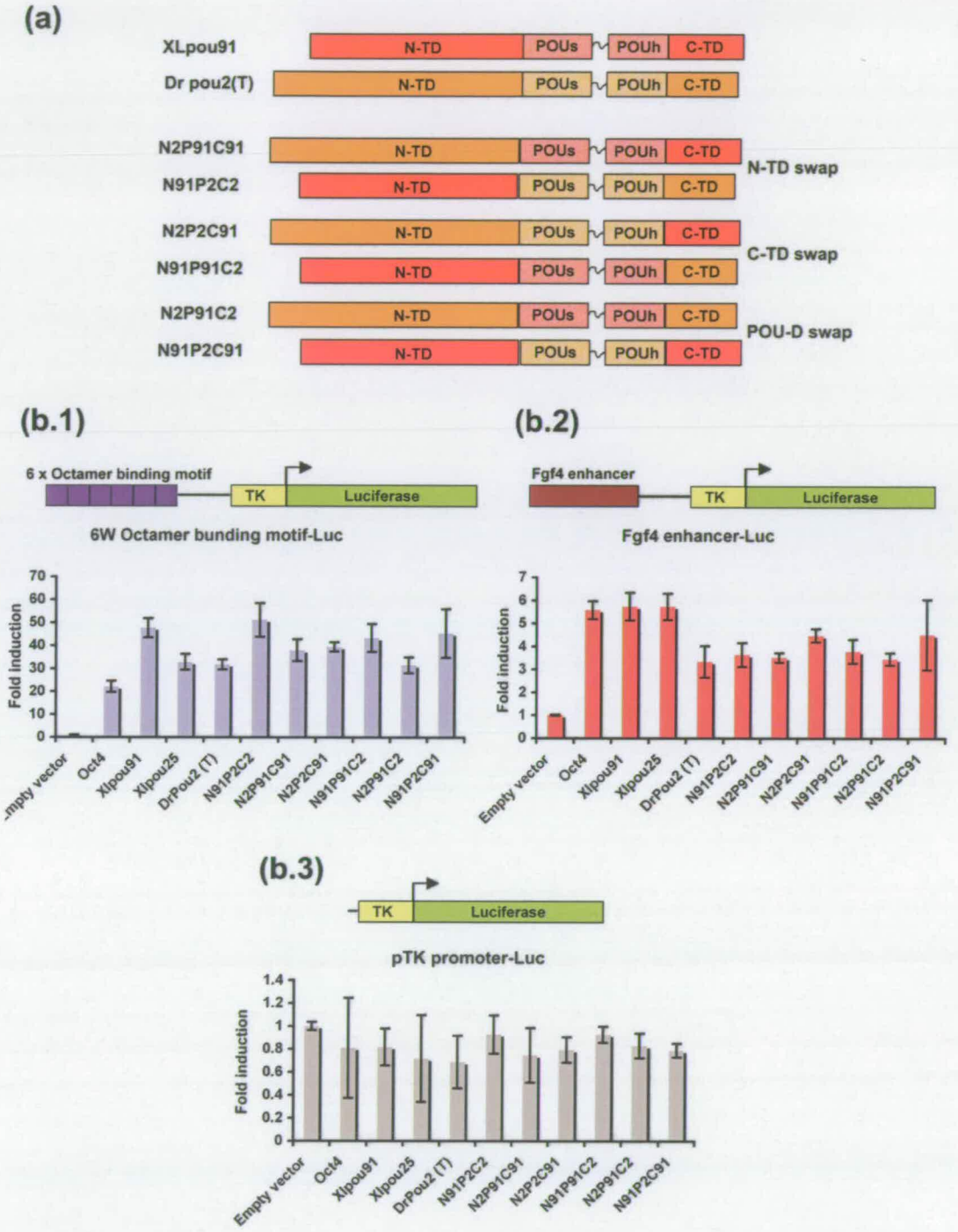


Figure 4.3-Cloning and transcriptional activity of Xlpou91-DrPou2(T) Chimeric proteins

(a) Schematic representation of Xlpou91-DrPou2(T) chimeric proteins: chimeric proteins were generated by swapping either the amino acid terminal (N-TD), the carboxyl terminal (C-TD) or the POU domain of Xlpou91 with DrPou2(T) domains. **(b)** Xlpou91-DrPou2(T) chimeric proteins activate transcription in a specific manner from two luciferase reporters: one containing multiple copies of the octamer binding motif **(b.1)** and the other one containing the Fgf4 enhancer **(b.2).** A luciferase reporter containing the thymidine kinase (TK) promoter was used to test the specificity of the chimeric proteins transcriptional activity **(b.3).**
The indicated cDNAs were cotransfected with luciferase reporters. Fold induction represents the increase in transcription compared with the empty vector control. Data represent the mean value of two independent experiments.

4.2.4. Measuring the ability of the different chimeric proteins to rescue short-term self-renewal: the generation of ES cell-like alkaline phosphatase positive colonies

As explained previously, the ability of the different chimeric proteins to rescue ES cell self-renewal was tested using the complementation assay designed by Niwa and colleagues (Niwa et al., 2002). The different chimeric proteins were introduced in the ZHBTc4 ES cells by random integration. To control for transfection efficiency, each transfection was divided in half with one half grown in the absence of Tc and the other half grown in presence of Tc. Following transfection, puromycin selection was applied for 9 days, and the resulting colonies were stained for alkaline phosphatase (AP) activity. The rescue index of each chimeric protein was then calculated by dividing the number of AP positive colonies in the presence of Tc by the number of AP positive colonies in the absence of Tc. The values of these rescue indices are normalised to that obtained from Oct4, and are shown in Figure 4.4 and 4.5.Figure 4.4b1 and 4.5b1 shows morphology of representative colonies rescued by various PouV proteins in the presence and absence of Oct4 regulatable transgene expression. As reported in Morrison and Brickman, 2006, colonies rescued with Xlpou91 protein exhibit similar morphology to those rescued with Oct4, whereas Xlpou25 rescued colonies are both small and differentiated. In the following sections, the phenotypes of colonies generated by each of the different chimeric proteins will be discussed. Summary of the ability of the different Xlpou91/Xlpou25, and Xlpou91/ DrPou2 (T) chimaeric proteins to activate transcriptions and to rescue ES cell self-renewal is listed in table 4.1 and able 4.2 respectively.

4.2.4.1. N-terminal domain of Xlpou91 has an inhibitory effect on the rate of ES cell self-renewal

Figure 4.4b2 shows the morphology of colonies rescued by the chimeric proteins N25P91C91 and N91P25C25. Colonies rescued with N91P25C25 protein have similar morphology to colonies rescued by Xlpou25 protein, while morphology of N25P91C91 rescued colonies is closer to those rescued with Xlpou91 with a slight increase in differentiated cells around the periphery.

Table 4.1- Summary of the ability of the different Xlpou91 and Xlpou25 chimaeric proteins to activate transcriptions and to rescue ES cell self-renewal.

The transcriptional activity of the different Xlpou91-Xlpou25 chimaeric proteins is assessed based on their ability to activate transcription from luciferase reporters harbouring Oct4 dependent sequences either from the Fgf4 enhancer or 6 reiterated copies of the Octamer binding motif. The ability of the different Xlpou91-Xlpou25 chimaeric proteins to rescue ES cell self-renewal is measured based on the morphology of the rescued colonies and the number of the rescued colonies represented as the rescue index. N: N-terminus, C:C-terminus, P:POU domain, 91: Xlpou91, 25: Xlpou25. Rescue index value is the mean from three independent experiments.

	Chimeric proteins	**Transcriptional activity**		**Morphology of rescued colonies**	**Rescue index**
		6W Octamer binding motif	Fgf4 enhancer		
	Oct4	+ +	+ + +	Typical undifferentiated ES cell colonies	1
	Xlpou91	+ + + +	+ + +	Typical undifferentiated ES cell colonies	1.01
	Xlpou25	+ + +	+ + +	Very small and differentiated colonies	0.47
	N91P25C25	+ + +	+ + +	Small and differentiated colonies, similar to Xlpou25 rescued colonies	0.53
	N25P91C91	+ + +	+ + +	Undifferentiated colonies similar to Xlpou91 rescued colonies, with slight increase in differentiated cells around the periphery	2.41
	N25P25C91	+ + +	+ + +	Mostly differentiated colonies that are bigger than Xlpou25 rescued colonies	1.10
	N91P91C25	+ + + +	+ + +	Undifferentiated ES cell colonies that are slightly smaller than Xlpou91 rescued colonies	0.38
	N25P91C25	+ + +	+ + +	Undifferentiated colonies slightly smaller than Xlpou91 rescued colonies, with slight increase in differentiated cells around the	0.67
	N91P25C91	+ + +	+ + +	Mostly differentiated colonies that are slightly bigger than Xlpou25 rescued colonies	0.91

Table 4.2- Summary of the ability of the different Xlpou91 and DrPou2(T) chimaeric proteins to activate transcriptions and to rescue ES cell self-renewal.

The transcriptional activity of the different Xlpou91-DrPou2(T) chimaeric proteins is assessed based on their ability to activate transcription from luciferase reporters harbouring Oct4 dependent sequences either from the Fgf4 enhancer or 6 reiterated copies of the Octamer binding motif. The ability of the different Xlpou91-DrPou2(T) chimaeric proteins to rescue ES cell self-renewal is measured based on the morphology of the rescued colonies and the number of the rescued colonies represented as the rescue index. N: N-terminus, C:C-terminus, P:POU domain, 91: Xlpou91, 2: DrPou2(T). Rescue index value is the mean from two independent experiments.

	Chimeric proteins	Transcriptional activity		Morphology of rescued colonies	Rescue index
		6W Octamer binding motif	Fgf4 enhancer		
	Oct4	+ +	+ + +	Typical undifferentiated ES cell colonies	1
	Xlpou91	+ + + +	+ + +	Typical undifferentiated ES cell colonies	1.02
	DrPou2(T)	+ + +	+ +	Small colonies with undifferentiated centre and differentiated cells around the periphery	0.33
	N91P2C2	+ + +	+ +	Small colonies similar to DrPou2(T) rescued colonies	0.07
	N2P91C91	+ + +	+ +	Undifferentiated colonies similar to Xlpou91 rescued colonies, with slight increase in differentiated cells around the periphery	1.85
	N2P2C91	+ + +	+ +	Colonies that are similar to DrPou2(T) rescued colonies, however they are bigger.	0.49
	N91P91C2	+ + +	+ +	Undifferentiated ES cell colonies that are slightly smaller than Xlpou91 rescued colonies	0.52
	N2P91C2	+ + +	+ +	Undifferentiated colonies similar to Xlpou91 rescued colonies, with slight increase in differentiated cells around the periphery	0.95
	N91P2C91	+ + +	+ +	Colonies that are similar to DrPou2(T) rescued colonies, however they are bigger	0.40

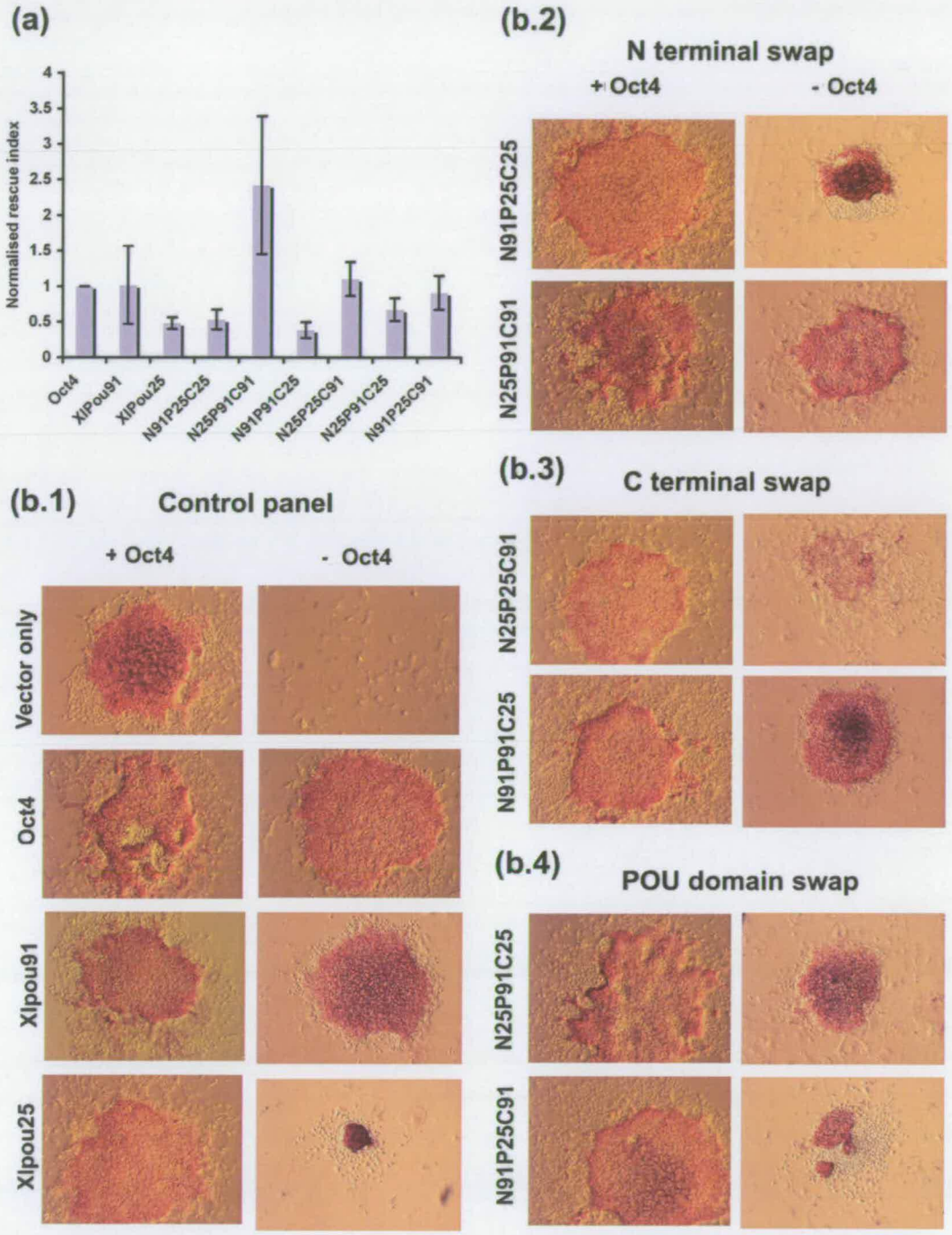


Figure 4.4-Xlpou91-Xlpou25 chimeric proteins differ in their ability to rescue ES cells self-renewal

$2x10^7$ ZHBTc4 ES cells were transfected with 100ug of linearized vector DNAs by electroporation, after 9 days of selection with puromycin in the presence or absence of tetracycline, the cells were stained with alkaline phosphatase (AP) and the number of AP positive stem cell colonies were scored, and the rescue index calculated.

(a) Rescue index of Xlpou91-Xlpou25 chimeric proteins. The Rescue index is calculated by dividing the number of ES cell colonies in the absence of Oct4 (presence of Tc) by the number of AP positive ES cell colonies prensent in the presence of Oct4 (absence of Tc). Rescue index values were normalised to Oct4 value. Data represents the mean values obtained from three independent experiments. **(b)** Morphology of colonies rescued by the Xlpou91-Xlpou25 chimeric proteins: morphology of representative colonies rescued by wild type PouV proteins **(b.1)**, chimeric protein generated by N terminal swap **(b.2)**, C terminal swap **(b.3)**, and POU domain swap **(b.4)**.

The lack of a transferable activity in the Xlpou91 amino terminus was also confirmed quantitatively as the N91P25C25 protein showed no increase in the rescue index when compared to full-length Xlpou25 (Figure 4.4a). Surprisingly, N25P91C91 protein showed a higher rescue index (around 2 folds) than that of either wild type Oct4 or Xlpou91. Together, the data suggest that the N-TD of Xlpou91 has little effect on the ES cell morphology, but appears to have a repressive effect on efficiency of self-renewal. These findings were confirmed by the results of the corresponding set of swaps between Xlpou91 and DrPou2(T) (Figure 4.5a and b2). In fact, N2P91C91 chimeric protein generated colonies that are morphologically similar to those rescued with Xlpou91, however, its rescue index was significantly greater than either Xlpou91 or Oct4.

Interestingly, N2P91C91 produced extremely differentiated colonies in the presence of the Tc regulatable *Oct4* transgene expression, a phenotype that might be caused by some unknown protein interactions with DrPou2 amino terminal.

4.2.4.2. C-terminal domain of Xlpou91 increases the efficiency of ES cell self-renewal

In order to test whether the C-TD of Xlpou91 protein has the ability to improve the performance of both Xlpou25 and DrPou2(T) proteins, we generated the chimeric protein N25P25C91 and N2P2C91 by swapping the carboxyl terminal of Xlpou25 and DrPou2 proteins respectively with the C terminal domain of Xlpou91 protein, and tested their ability to rescue ES cell self-renewal. As controls, N91P91C25, and N91P91C2 chimeric proteins were made by substituting the C terminal domain of Xlpou91 protein with the C terminal from Xlpou25 or DrPou2 respectively.

Figure 4.4b3 shows the morphologies of rescued colonies by N25P25C91 and N91P91C25 chimeric proteins. Although, N25P25C91 protein produces differentiated colonies that are bigger than the ones produced by Xlpou25, it has an Oct4-like rescue index (Figure 4.4a). A reciprocal result was obtained with N91P91C25, this protein was able to produce colonies with a normal ES-like morphology, but had a reduced Xlpou25-like rescue index (Figure 4.4a and b3). Similar results were obtained with N2P2C91, and N91P91C2 chimeric proteins.

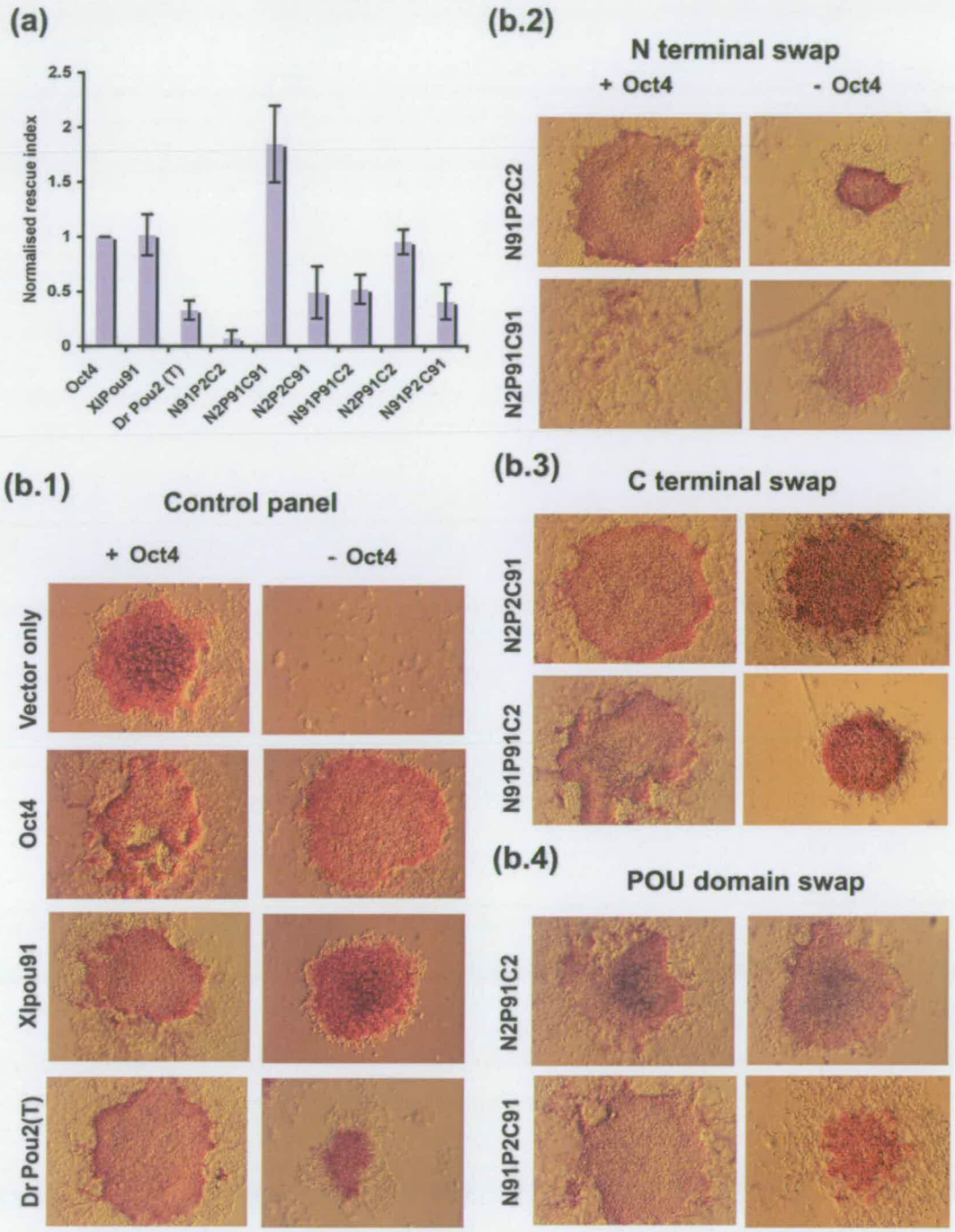


Figure 4.5-Xlpou91-DrPou2(T) and Xlpou91-Xlpou25 chimeric proteins show similar ability to rescue ES cells self-renewal

(a) Rescue index for Xlpou91-DrPou2(T) chimeric proteins. Rescue index values were normalised to Oct4 value. Data represents the mean values obtained from two independent experiments. **(b)** Morphology of colonies rescued by the Xlpou91-DrPou2(T) chimeric proteins: morphology of representative colonies rescued by wild type PouV proteins **(b.1)**, chimeric proteins generated by N terminal swap **(b.2)**, C terminal swap **(b.3)**, and POU domain swap **(b.4)**.

Figure 4.5b3 shows that N2P2C91 generated colonies that are bigger than those rescued with DrPou2 protein but more differentiated than typical ES cell colonies and had a greater rescue index than that achieved by DrPou2. Moreover, N91P91C2 protein also recapitulates the effect of N91P91C25 protein as it produces colonies that are morphologically very similar to those rescued with Xlpou91 albeit slightly smaller and had a lower rescue index than Xlpou91 or Oct4 (Figure 4.5a and b3).
Therefore, these data suggest that Xlpou91 C terminal influences positively both the size and the number of rescued colonies.

4.2.4.3. POU DNA binding domain from Xlpou91 protein is associated with the undifferentiated phenotype of ES cells

Chimeric proteins generated by substituting the POU DNA binding domain of Xlpou25 and DrPou2 with that of Xlpou91 protein and vice-versa, were used to examine the role of the POU DNA binding domain in the rescue of ES cell self-renewal.
While not always achieving quantitative rescue as determined by the rescue index, the POU domain appears to rescue ES cell morphologies most explicitly. Colonies rescued with N25P91C25 (Figure 4.4b4) or N2P91C2 proteins (Figure 4.5b4) exhibit very similar morphology to those rescued with Xlpou91, but show a slight increase in the number of differentiated cells surrounding the colonies. On the other hand, N91P25C91, and N91P2C91 chimeric proteins produced colonies that are slightly bigger than those obtained with Xlpou25 and DrPou2 respectively, but still show differentiated morphologies (Figure 4.4b4) and (Figure 4.5b4).
These results indicate that the POU DNA binding domain of Xlpou91 influences hugely the morphology of the cells, and determines the undifferentiated phenotype of ES cells. In addition to this effect on morphology, Xlpou91 POU DNA binding domain seems to have a positive effect on the number of rescued colonies as Figures 4.4a and Figure 4.5a demonstrate that both rescue indices of N25P91C25 and N2P91C2 proteins are higher than those from Xlpou25 and DrPou2 respectively.

4.2.5. Long-term self-renewal: the generation of clonal cell lines from the rescued colonies

In order to confirm the findings regarding the ability of the various chimeric proteins to support ES cell self-renewal, colonies rescued by each chimeric protein were examined for their ability to generate cell lines that are able to propagate in self-renewing conditions. Flag tagged chimeric proteins were used to generate these cell lines. Three copies of the Flag tag (3xFlag) were fused to the 5' amino terminal of each chimeric protein (Figure 4.6a). The triple Flag tag was used for two main reasons, the lack of antibodies against any of the PouV proteins and the generation of reagents that will be useful for downstream biochemical analysis.

4.2.5.1. The addition of the Flag tag to the chimeric proteins does not alter their function

In order to establish whether or not fusing the Flag tag to the amino terminal of the chimeric proteins will affect their function, both the abilities of these Flag tagged chimeric proteins to activate transcription and sustain ES cell self-renewal were examined. Figure 4.6 shows that no differences were observed in the ability of chimeric proteins with or without the Flag tag to activate transcription from the octamer-binding motif luciferase reporter (Figure 4.6b1), or the *Fgf4* enhancer luciferase reporter (Figure 4.6b2) and that none of these proteins had any activity on a reporter lacking octamer binding sites (Figure 4.6b3). Moreover, their ability to rescue ES cell self-renewal was almost identical to that obtained by the untagged proteins when assayed for either ES cell morphology or rescue index (Figure 4.7). These observations both support my conclusions with respect to the different domain swaps and indicate that cell lines derived from these tagged proteins will be important tools for the analysis of domain specific activities of these PouV proteins.

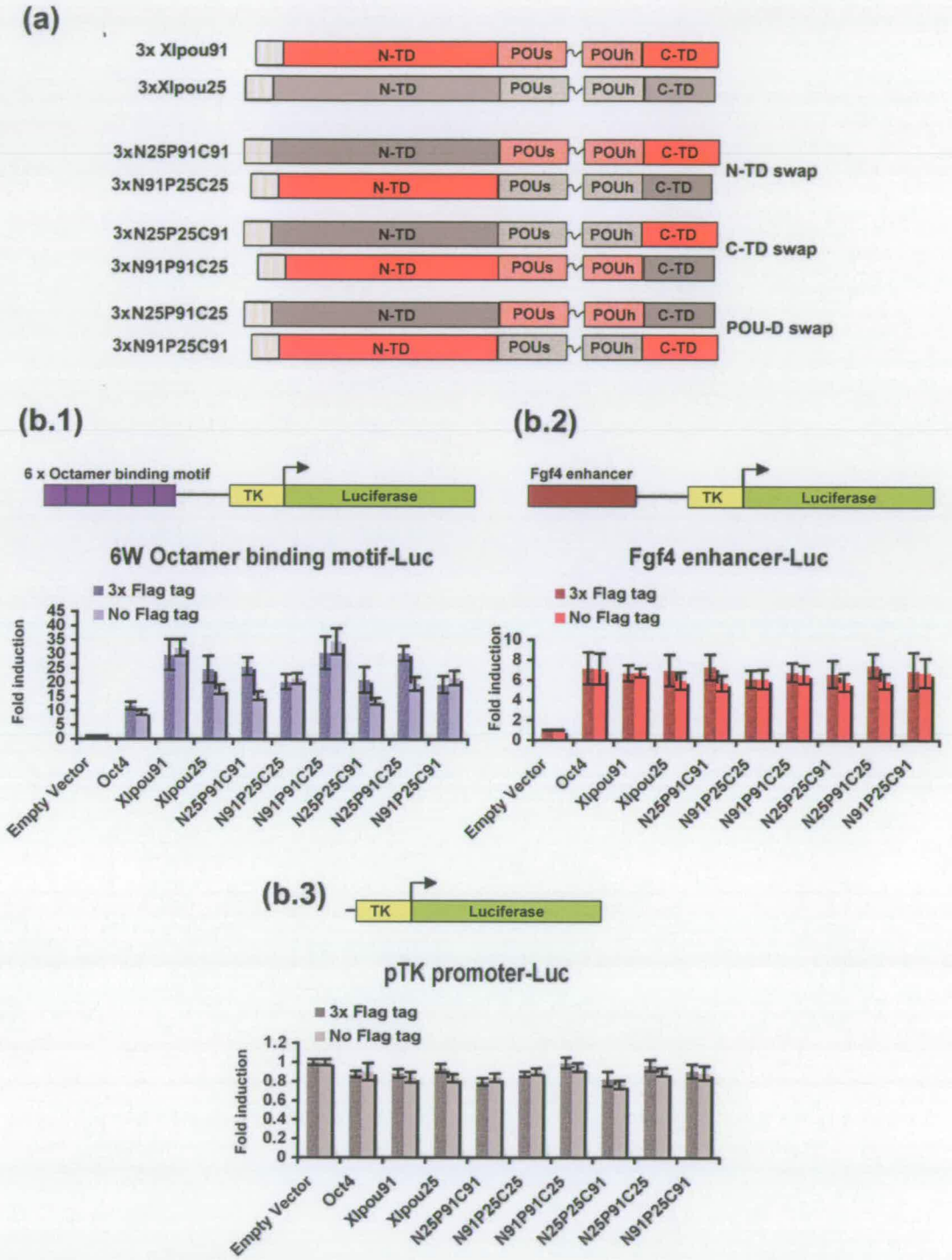


Figure 4.6–Cloning and transcriptional activity of Flag tagged Xlpou91-Xlpou25 Chimeric proteins

(a) Schematic representation of 3xFlagXlpou91-Xlpou25 chimeric proteins: chimeric proteins were generated by swapping either the amino acid terminal (N-TD), the carboxyl terminal (C-TD) or the POU domain of Xlpou91 with Xlpou25 domains. Three copies of the flag tag were fused to the amino terminal of the chimeric proteins. **(b)** Xlpou91-Xlpou25 chimeric proteins with or without the Flag tag activate transcription at similar levels in a specific manner from two luciferase reporters: one containing multiple copies of the octamer binding motif **(b.1)** and the other one containing the Fgf4 enhancer **(b.2).** A luciferase reporter containing the thymidine kinase (TK) promoter was used to test the specificity of the chimeric proteins transcriptional activity **(b.3).**

The indicated cDNAs were cotransfected with luciferase reporters. Fold induction represents the increase in transcription compared with the empty vector control. Data represent the mean value of two independent experiments.

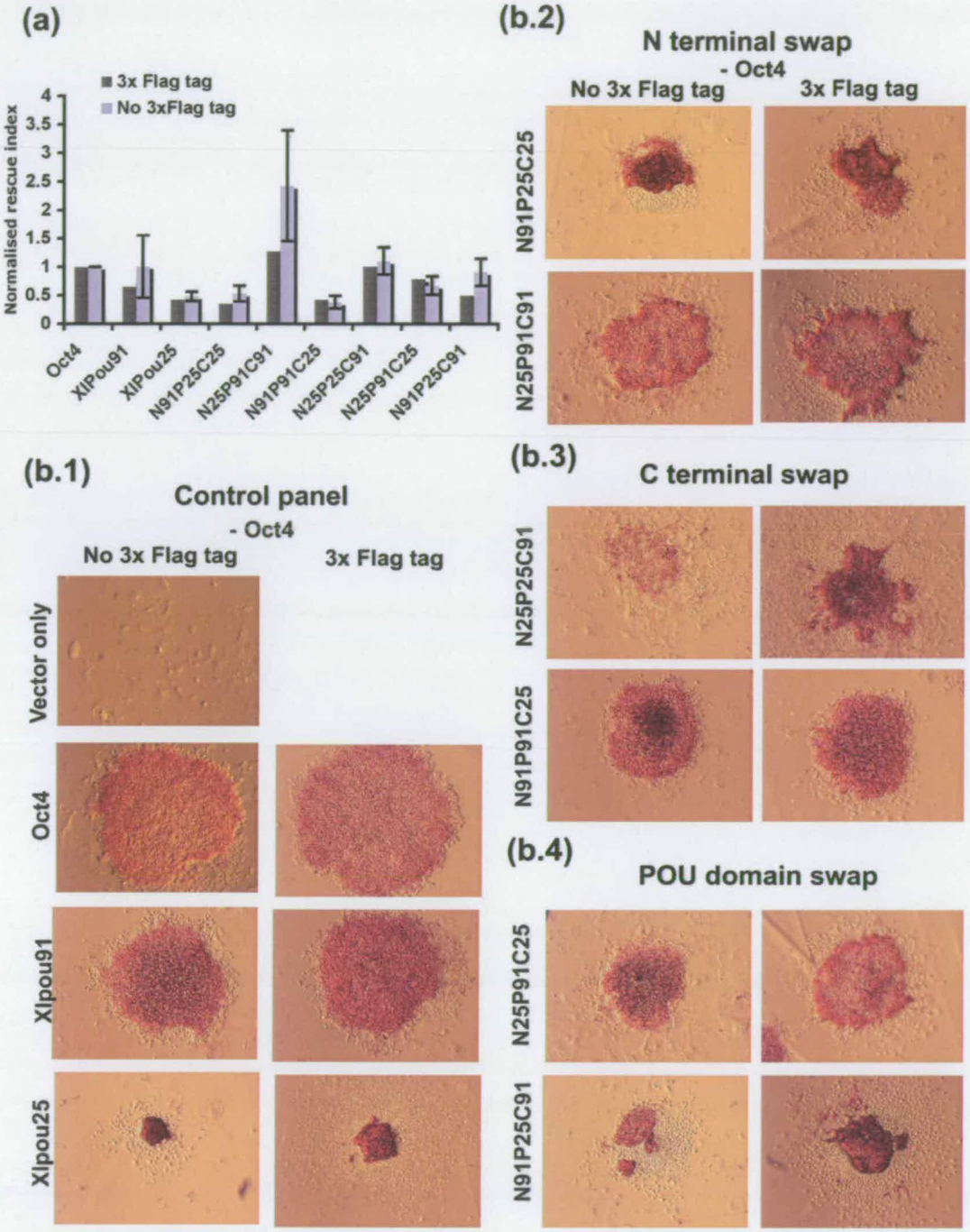


Figure 4.7-The addition of the Flag tag to the Xlpou91-Xlpou25 chimeric proteins does not affect their ability to rescue ES cell self-renewal

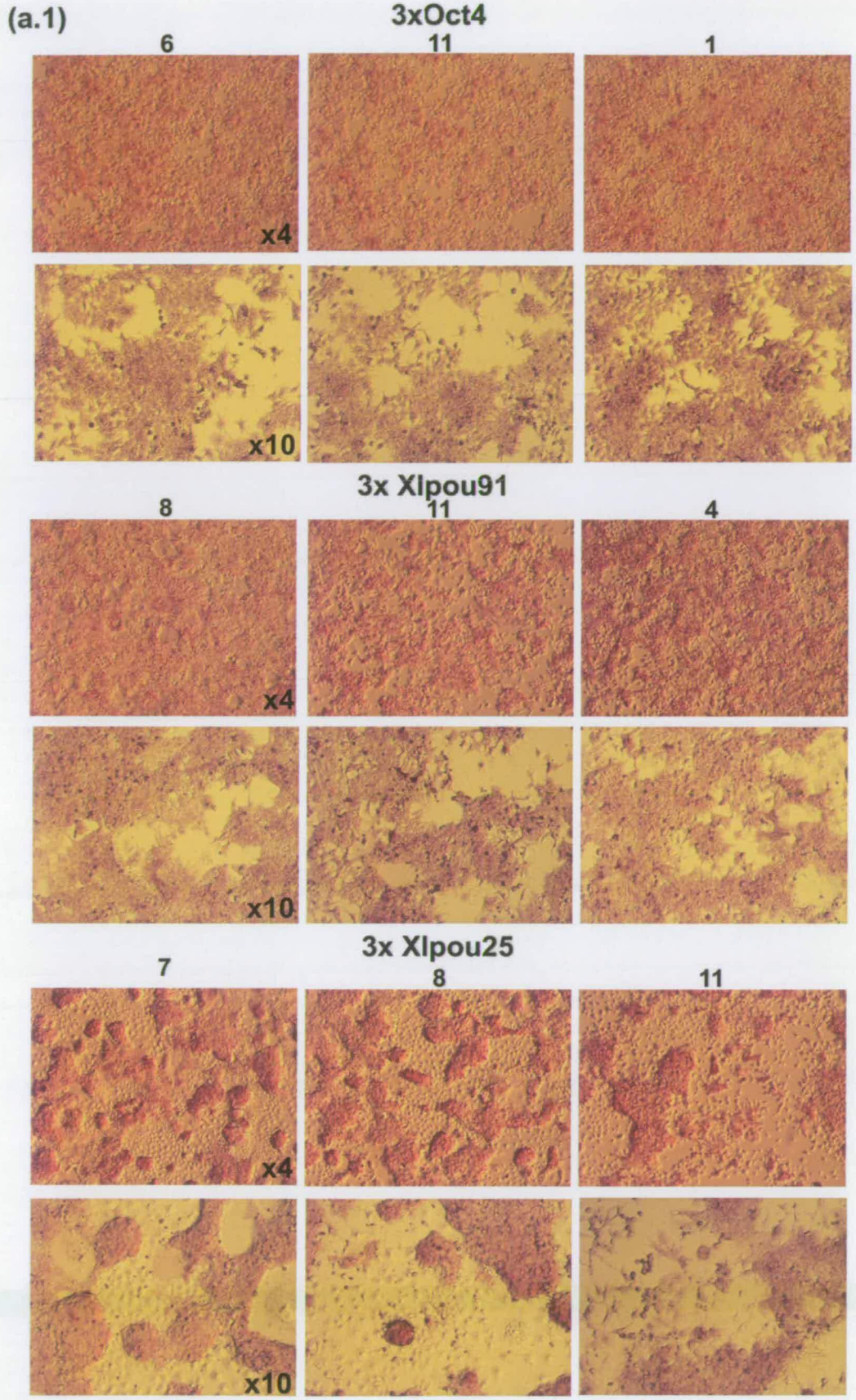
(a) Xlpou91-Xlpou25 chimeric proteins with or without Flag tag have similar rescue indices. Rescue index values of chimeric proteins with or without Flag tag were normalised to 3xFlagOct4 and Oct4 values respectively. Rescue index of Flag tagged proteins is from one experiment. **(b)** Colonies rescued by Xlpou91-Xlpou25 chimeric proteins with or without Flag tag have similar morphologies. Morphology of representative colonies rescued by wild type PouV proteins **(b.1)**, chimeric proteins generated by N terminal swap **(b.2)**, C terminal swap **(b.3)**, and POU domain swap **(b.4)**.

4.2.5.2. Derivation of clonal cell lines from the Flag tagged Xlpou91-Xlpou25 chimeric proteins

Clonal cell lines were derived by expanding puromycin resistant colonies generated from electroporation of the ZHBTc4 ES cells with Flag tagged Xlpou91-Xlpou25 chimeric cDNAs in the presence of Tc (absence of Tc regulatable *Oct4* transgene expression). The rescued colonies were propagated for many passages under self-renewing conditions, and stained for Alkaline phosphatase (AP) activity. Figure 4.8 shows alkaline phosphatase staining for each cell line with images of the morphology at both high (x10) and low magnifications (x4).

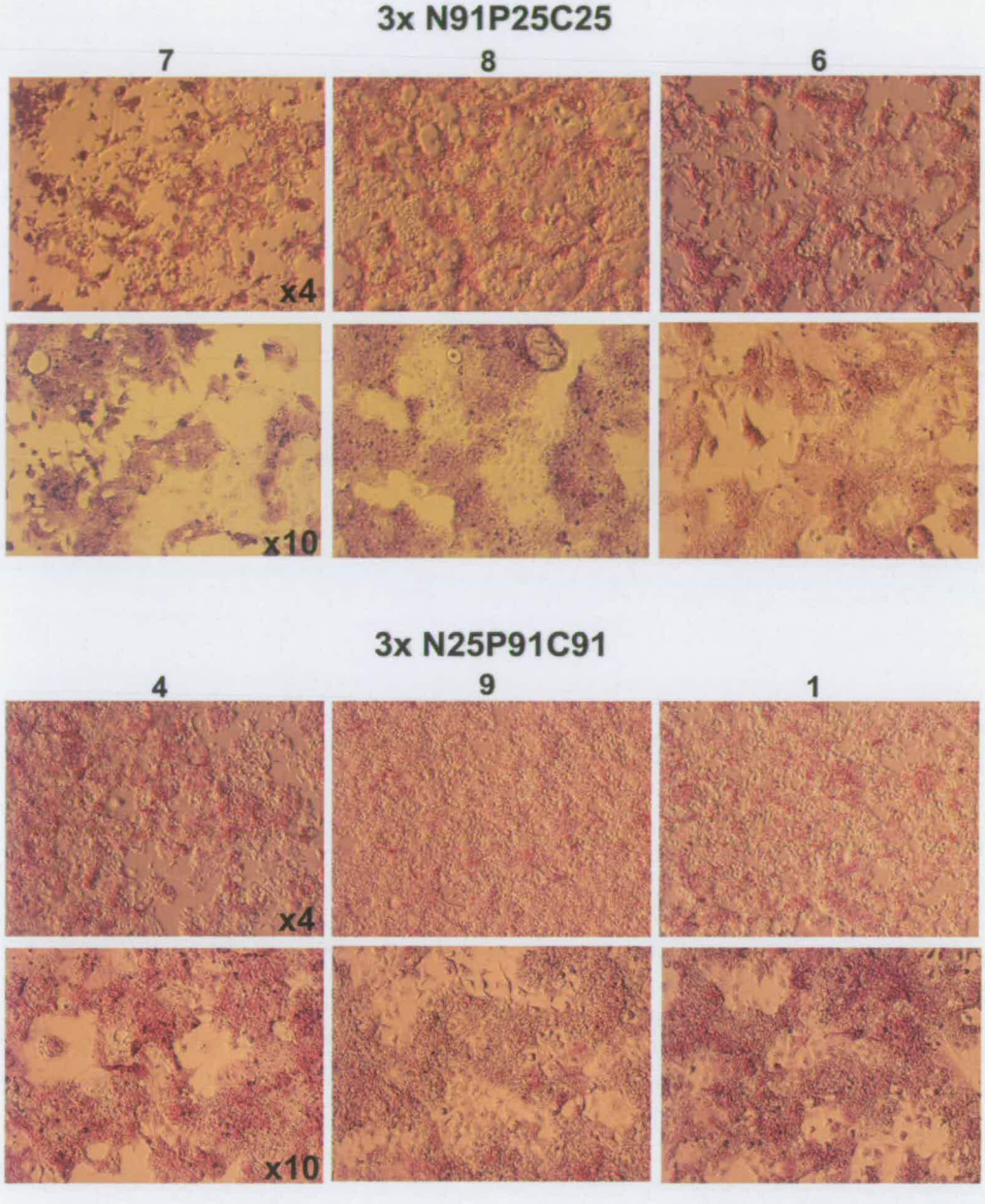
Three clonal cell lines for each chimeric protein were examined for their morphology. Numbers on the top of each photograph represent the name of the clone used for the derivation of the cell line.

The morphology of cell lines derived from colonies rescued with Flag tagged Oct4 or Xlpou91 proteins resembles the morphology of the parental ZHBTc4 and E14Tg2a cell lines, they are homogenous undifferentiated cells that are positive for Alkaline phosphatase staining (Figure 4.8a1). This contrasts the morphology of cell lines rescued with the Flag tagged Xlpou25 protein. The phenotype shown in Figure 4.8a1 recapitulates that described by Morrison and Brickman 2006 and features a layer of differentiated cells with AP positive clumps of cells that appear to grow as colonies. As with the original Xlpou25 supported cell lines (Morrison and Brickman, 2006), these are very slow growing. Cell lines stably expressing chimeric proteins that have only the C-TD or N-TD of Xlpou91 (N91P25C25, or N25P25C91) exhibit some rescue, but they are mainly differentiated cells that grow slowly (they take longer to reach confluence than Xlpou91 or Oct4 cell line when plated at the same density) (Figure 4.8a2, and a3). However, chimeric proteins that possess the Xlpou91 POU domain (N25P91C91 or N25P91C25) are predominantly undifferentiated. Furthermore, the importance of Xlpou91 POU domain in maintaining the undifferentiated phenotype of cells is confirmed by the phenotypes of cells maintained by the chimeric protein N91P25C91, it only supports the differentiated phenotype observed in Xlpou25 rescued cell lines (Figure 4.8a4).


(a.1)
3xOct4
6
11
1
x4
x10
3x Xlpou91
8
11
4
x4
x10
3x Xlpou25
7
8
11
x4
x10

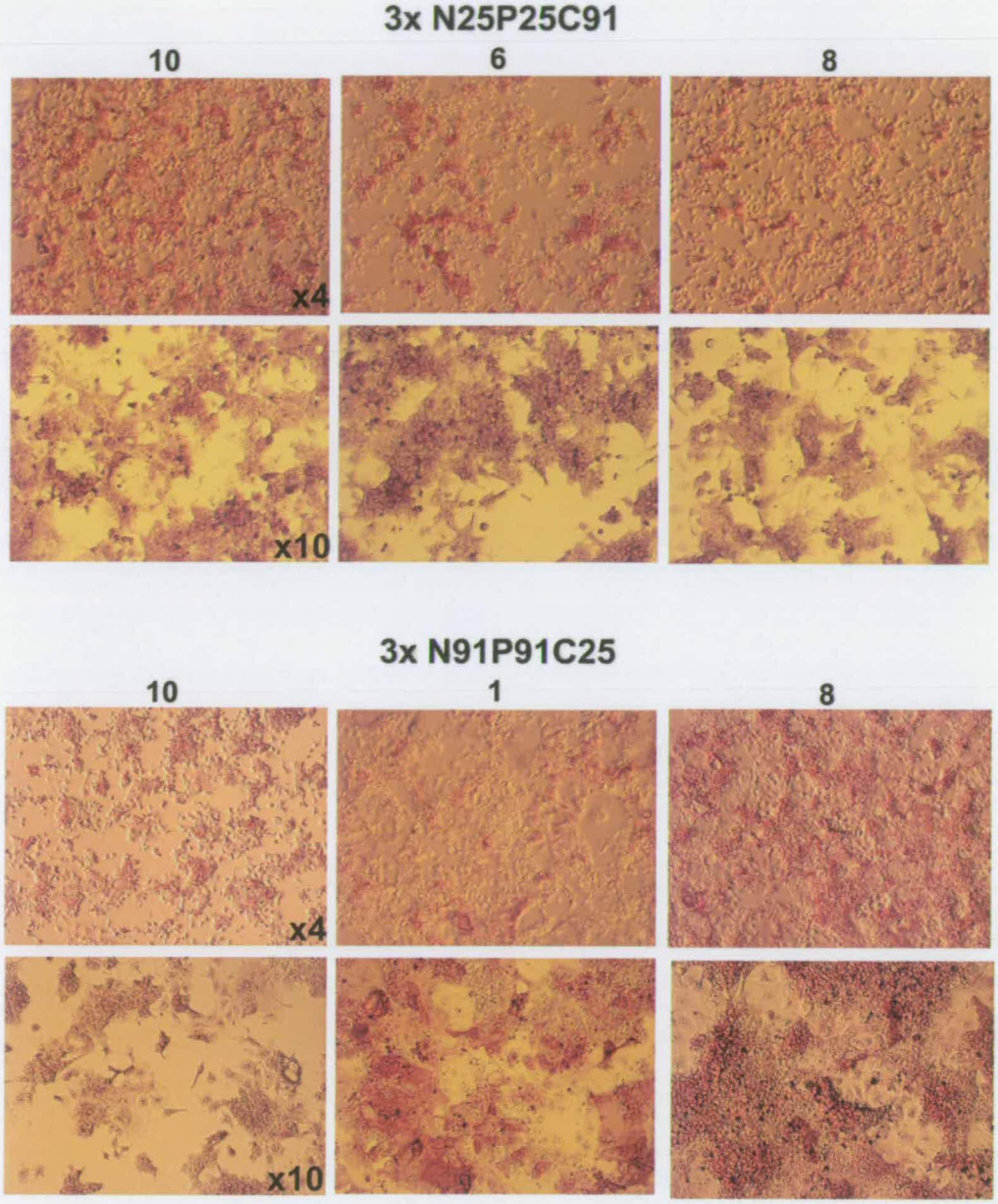
(a.2)

N-Terminal swap



(a.3)

C-Terminal swap



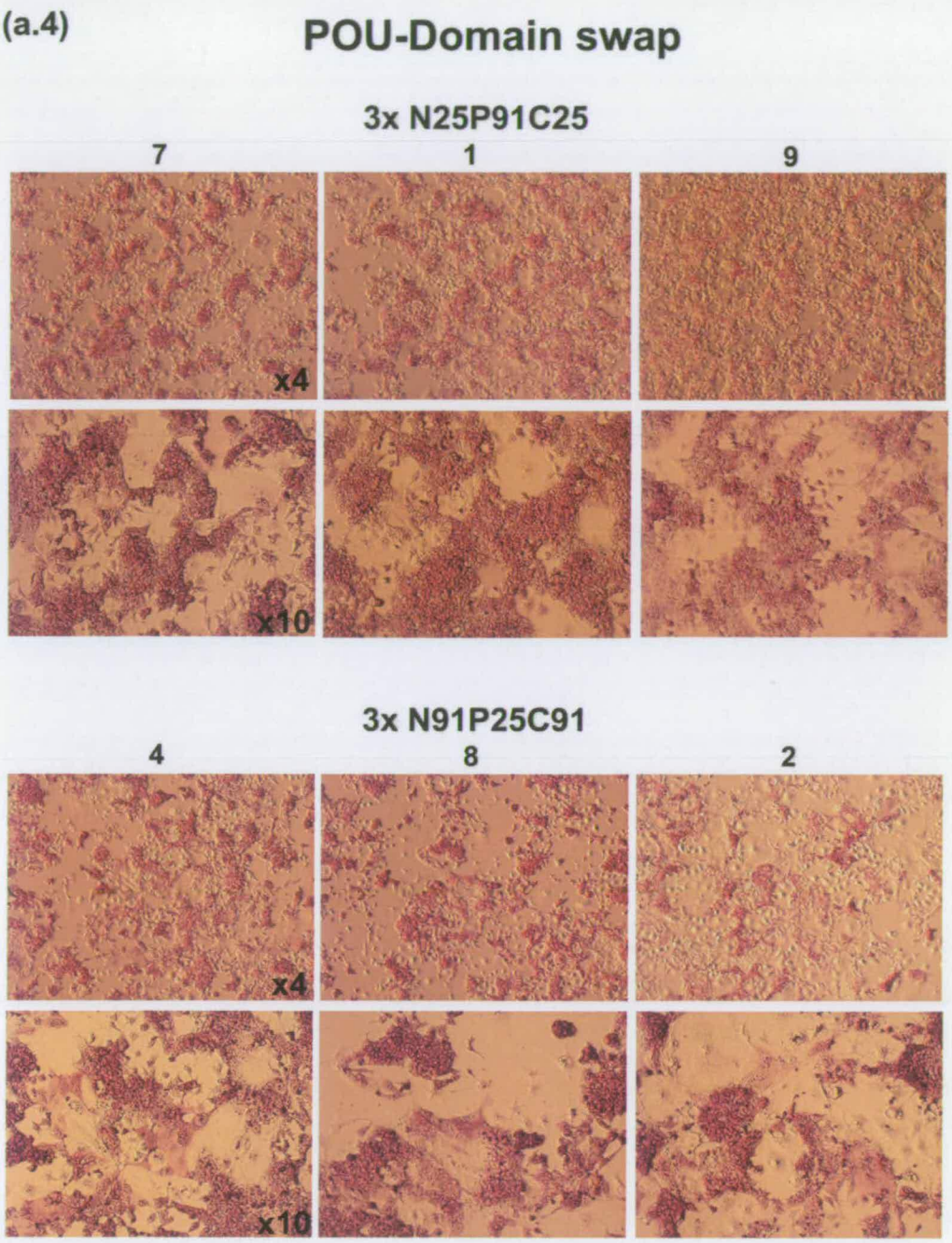


Figure 4.8-Derivation of clonal cell lines rescued with 3xFlagXlpou91-Xlpou25 chimeric proteins

Following the electroporation of construct encoding the Flag tagged Xlpou91-Xlpou25 chimeric proteins in the ZHBTc4 cells, puromycin selection was applied for 9days in the presence of tetracycline. The rescued colonies were propagated for many passages under self-renewing conditions in the presence of Tc, and stained for Alkaline phosphatase (AP) activity.

Morphology of cell lines derived from wild type PouV proteins **(a.1)** chimeric proteins generated by N terminal swap **(a.2)**, C terminal swap **(a.3)**, and POU domain swap **(a.4)** is shown at both high (x10) and low magnification (x4).
Three clonal cell lines for each chimeric protein were examined for their morphology. Numbers represent the names of the cell lines.

Although, the morphologies of these derived cell lines is in accordance with the different self-renewal rescue abilities of the various chimeric proteins, more extensive characterization of these cell lines is needed. However, as a first step we wanted to ensure that the derived cell lines express similar levels of the chimeric proteins. To this end, western blot analysis with anti-Flag antibody was carried out on lysates of cell lines that had been growing in culture under self-renewing conditions and puromycin selection (Figure 4.9). Lysates from the ZHBTc4 and E14Tg2a cell lines were used as negative controls. The cell lysates from the other cell lines showed specific bands that correspond to the predicted size of each chimeric protein. 3xOct4 cell line produced a 41.23 KDa, 3xXlpou91: 52.32 KDa, 3xXlpou25: 52.87 KDa , 3xN91P25C25: 51.51 KDa, 3xN25P91C91: 53.68 KDa, 3xN25P25C91: 53.6 KDa, 3xN91P91C25: 51.58 KDa, 3xN25P91C25: 52.94 KDa, 3xN91P25C91: 52.25 KDa.

Equal amount of protein was used for each sample as confirmed with β-actin levels.

Numbers represents the names of the cell lines used in the analysis.

In general, the levels of protein were similar, and where their levels varied, there was no correlation between the amount of the protein produced in the cell lines and the rescue ability of the proteins.

As there appears no significant difference in expression level or transcriptional activity of these chimeric proteins in transient transfection, I was interested in investigating the contribution of the C-TD and POU domain to self-renewal and explore potential mechanism(s). While detailed mechanistic studies are beyond the scope of this thesis, we begin by considering each domain in detail.

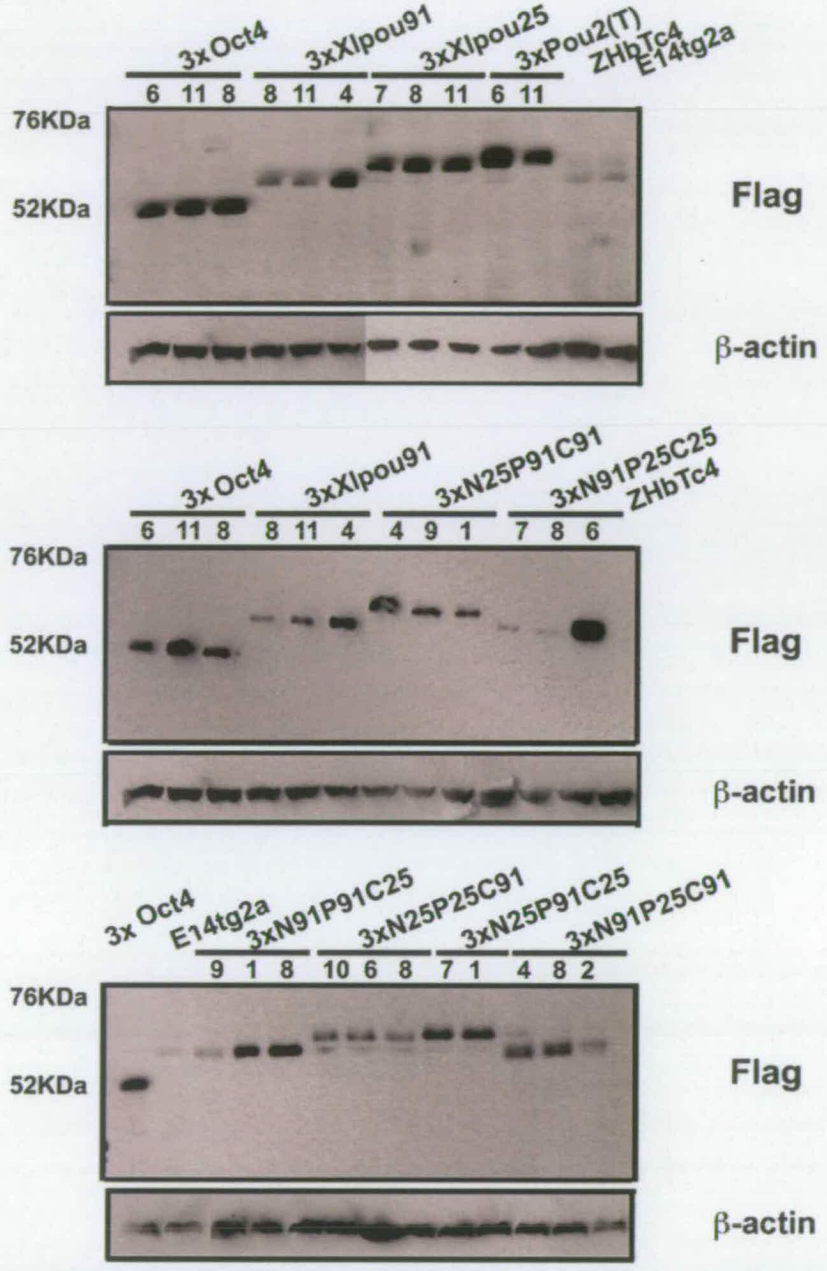


Figure 4.9-Western blot analysis of 3xFlag-Xlpou91-Xlpou25 chimeric proteins expression levels in the rescued cell lines

Cell lines derived from rescued colonies by the 3xFlag Xlpou91-Xlpou25 chimeric proteins were cultured under self-renewing conditions, then lysed for western blot analysis with anti Flag , and anti β-actin antibodies.
Equal amount of protein was used for each sample.
Numbers represents the names of the cell lines

4.2.6. Investigating the importance of Xlpou91 C-terminal domain

4.2.6.1. Deletion of the N- and C-terminal domains from Xlpou91 and Oct4 proteins reveals the important role of the C-TD in increasing the efficiency of ES cell self-renewal rescue

The data presented above (Figure 4.4 and 4.5), indicated a role for the C-TD of Xlpou91 in augmenting self-renewal efficiency. We wanted to directly test the significance of this domain and investigate its role in both Xlpou91 and Oct4 proteins. To achieve this, we made a series of deletion mutants for both Oct4 and Xlpou91 proteins lacking either the N terminal domain or the C terminal domain (Figure 4.10a), and tested their ability to activate transcription and rescue ES cell self-renewal. Table 4.3 summarises the ability of the different deletion mutants for both Oct4 and Xlpou91 proteins to activate transcriptions and to rescue ES cell self-renewal.

Figure 4.10b1 shows that Oct4 and Xlpou91 proteins lacking either the N-TD or the C-TD are all impaired with respect to the activation of the luciferase reporter containing six copies of the octamer-binding motif. Interestingly, this defect in transcription activity appears most pronounced in proteins lacking the N-TD. On the other hand, all these mutant proteins appear equal to the full-length proteins in their ability to activate transcription from the *Fgf4* enhancer reporter (Figure 4.10b2). As *Fgf4* enhancer also binds Sox2, we presume that the activation function of Sox2 maybe masking any defects in the transactivation ability of its PouV protein partner. The specificity of the transcriptional activity of the deletion mutant proteins was confirmed as they were unable to stimulate transcription from the tk promoter luciferase reporter (Figure 4.10b3). Moreover, all Flag and non Flag tagged proteins had identical activity confirming that the Flag tag does not alter the activity of the proteins.

Figure 4.11b2 shows that both ΔN or ΔC Oct4 mutant proteins produced morphologically normal undifferentiated cells, but showed lower rescue indices in comparison to the full-length Oct4 protein (Figure 4.11a). This reduction in the rescue index was also reported previously by Hitoshi Niwa's laboratory (Niwa et al., 2002).

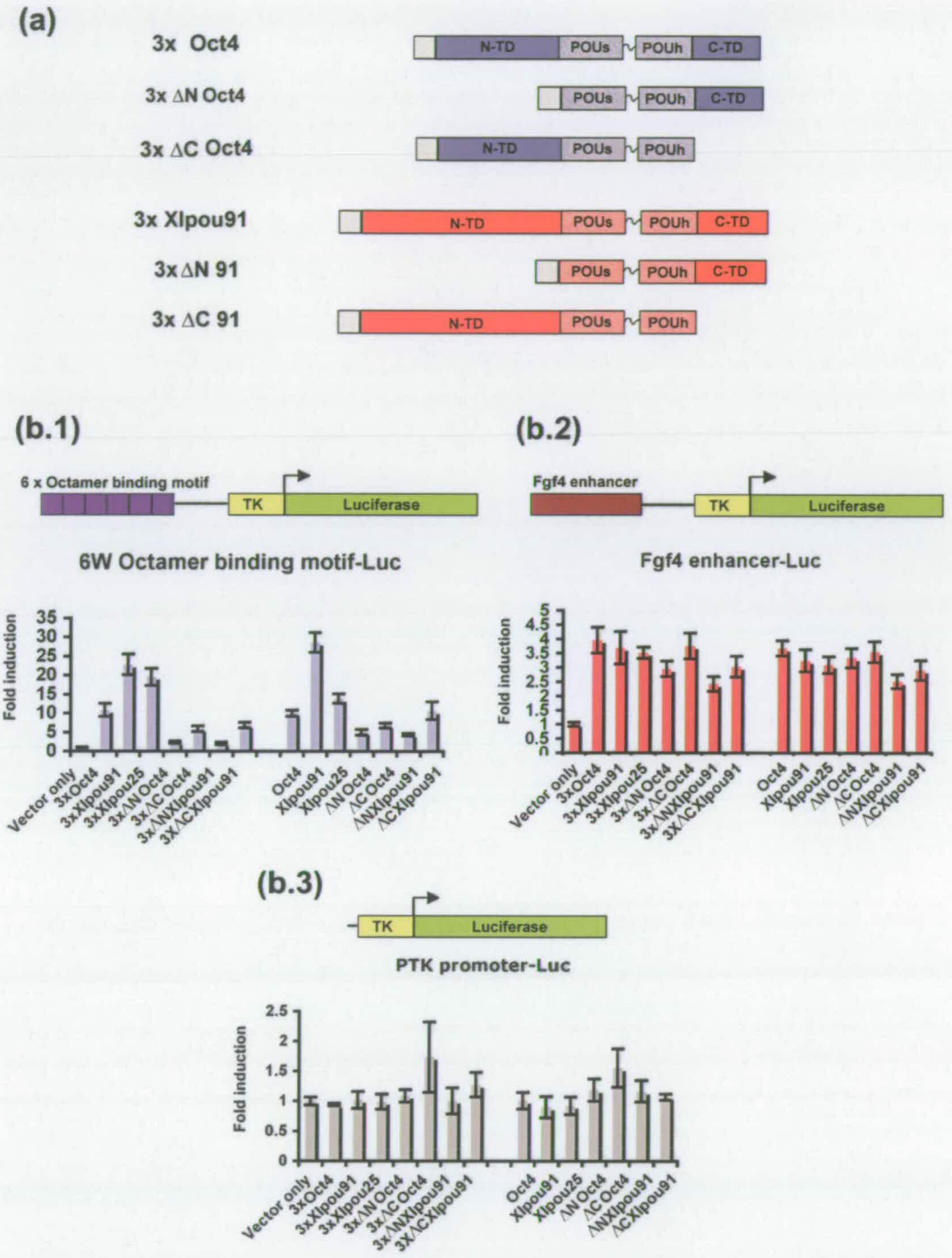


Figure 4.10 -Cloning and transcriptional activity of Flag tagged truncated Oct4 and Xlpou91 proteins

(a) Schematic representation of 3xFlag tagged truncated Oct4 and Xlpou91 proteins: either the N terminal or the C terminal of the proteins was deleted.

(b) Truncated Oct4 and Xlpou91 proteins with or without the Flag tag activate transcription at similar levels in a specific manner from two luciferase reporters: one contains multiple copies of the octamer binding motif **(b.1)** and the other contains the Fgf4 enhancer **(b.2).** A luciferase reporter containing the thymidine kinase (TK) promoter was used to test the specificity of the proteins transcriptional activity **(b.3).**

The indicated cDNAs were cotransfected with luciferase reporters. Fold induction represents the increase in transcription compared with the empty vector control. Data represent the mean value of two independent experiments.

Table 4.3- Investigating the importance of Xlpou91 C-terminal domain.

The transcriptional activity of the different deletion and DEF site mutant Oct4 and Xlpou91 constructs is assessed based on their ability to activate transcription from luciferase reporters harbouring Oct4 dependent sequences either from the Fgf4 enhancer or 6 reiterated copies of the Octamer binding motif. The ability of the different mutant Oct4 and Xlpou91 constructs to rescue ES cell self–renewal is measured based on the morphology of the rescued colonies and the number of the rescued colonies represented as the rescue index. 3x designated 3 copies of the flag tag. N: N-terminus, C:C-terminus, P:POU domain, 91: Xlpou91, 25: Xlpou25, Δ: deletion. Rescue index value is the mean from two independent experiments.

	Chimeric proteins	Transcriptional activity		Morphology of rescued colonies	Rescue index
		6W Octamer binding motif	Fgf4 enhancer		
Wild type proteins	3x Oct4	++	+++	Typical undifferentiated ES cell colonies	1
	3x Xlpou91	++++	+++	Typical undifferentiated ES cell colonies	0.65
	3x Xlpou25	+++	+++	Very small and differentiated colonies	0.45
N terminal deletion	3x ΔN Oct4	+	+++	Undifferentiated colonies similar to Oct4 rescued colonies, with slight increase in differentiated cells around the periphery	0.31
	3x ΔN 91	+	+++	Undifferentiated colonies similar to Xlpou91 rescued colonies, with slight increase in differentiated cells around the periphery	0.33
C terminal deletion	3x ΔC Oct4	Slightly more than +	+++	Undifferentiated colonies similar to Oct4 rescued colonies, with slight increase in differentiated cells around the periphery	0.23
	3x ΔC 91	Slightly more than +	+++	Highly differentiated colonies	0
DEF site mutants	3x Mutant Oct4	+++	+++	Undifferentiated ES cell colonies, identical to Oct4 rescued colonies	0.88
	3x Mutant 91	++++	+++	Undifferentiated ES cell colonies that are identical to Xlpou91 rescued colonies	0.46

The previous results contrast to those observed with Xlpou91, where deletion of the C-TD completely abrogated the ability of Xlpou91 to rescue ES cell self-renewal. In fact, only ΔN Xlpou91 protein was able to produce undifferentiated ES cell colonies (Figure 4.11b2 and b3). Although, deletion of the amino terminus of Xlpou91 did not have much impact on the morphology of the rescued colonies, it resulted in a reduction of the rescue index (Figure 4.11a), which suggests that the N-TD is required to achieve optimal rescue.

These results indicate that in combination with the POU domain both the N-TD and the C-TD of Oct4 protein possess the ability to maintain ES cell self-renewal, whereas in the Xlpou91 protein only the C terminal domain has this ability.

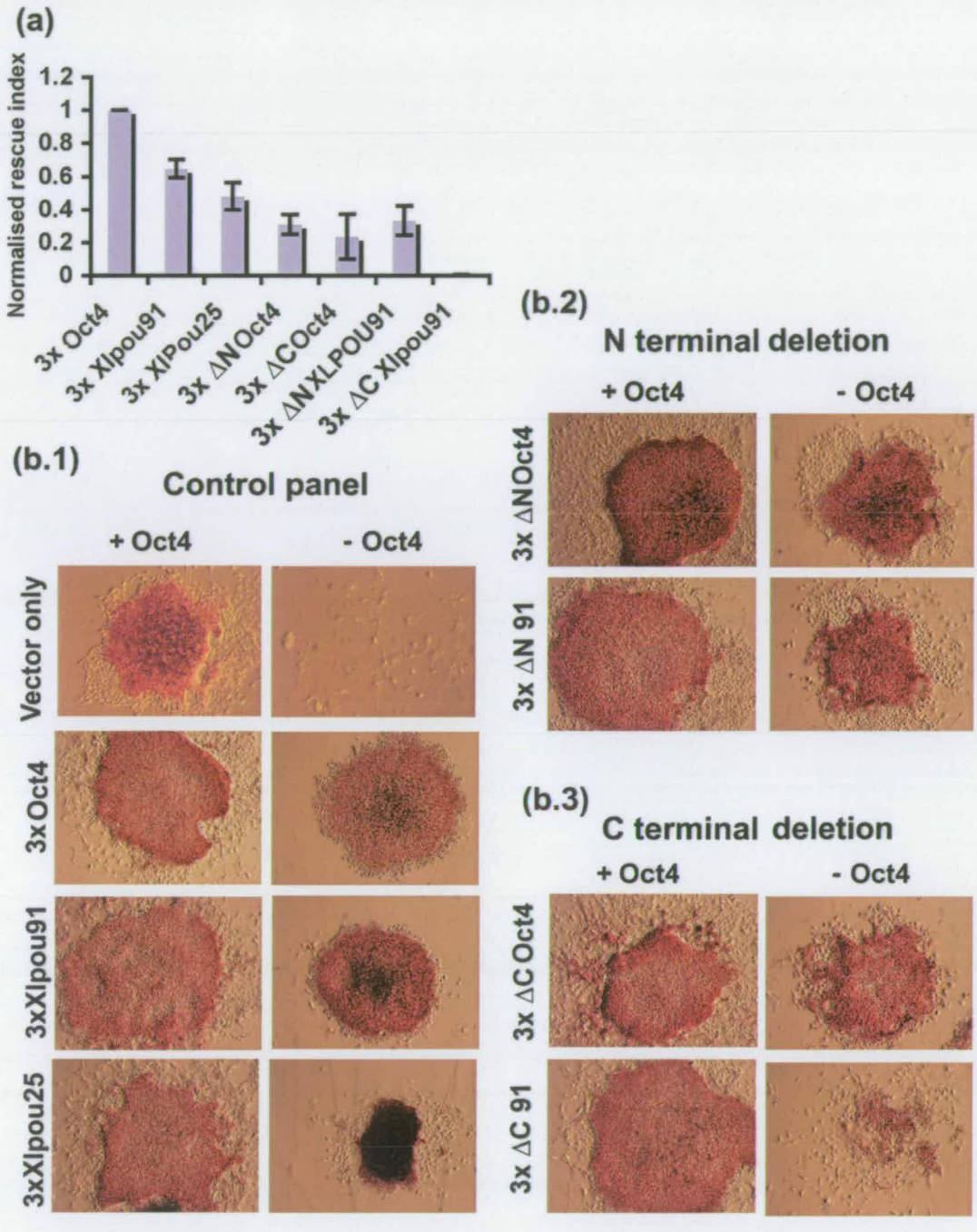


Figure 4.11-Xlpou91 protein missing the C terminal is unable to rescue ES cells self-renewal

(a) Rescue index of flag tagged truncated Oct4 and Xlpou91 proteins. The rescue index values were normalised to 3xFlagOct4 value. Data represents the mean values obtained from two independent experiments. **(b)** Morphology of colonies rescued by the Flag tagged truncated Oct4 and Xlpou91 proteins. Morphology of representative colonies rescued by full length PouV proteins **(b.1)**, proteins without N terminal **(b.2)**, or proteins without the C terminal **(b.3)**.

4.2.6.2. Protein alignment of PouV proteins reveals a possible functional DEF (docking site for ERK, $^{F}/_{Y}$ -X-$^{F}/_{Y}$ -P) site in the C terminal domain

As the Xlpou91 C-TD appears to convey an advantage to the other PouV proteins in their ability to rescue *Oct4* null phenotype, we looked for sequences in Oct4 and Xlpou91 C-TDs that may be responsible for conferring the high self-renewal rescue ability. We hypothesized that these elements exist only in the C-TD of proteins with high rescue ability, and not the other PouV proteins with diminished ability to rescue ES cells self-renewal. We therefore, aligned the amino acid sequences of the C-TD from different PouV proteins using Clustal W multiple sequence alignment software (Figure 4.12a).

The alignment uncovered a motif for a potential tyrosine kinase–Ras–mitogen-activated protein kinase (MAPK) docking site that was present only in mouse Oct4 (mOct4), human OCT4 (hOCT4), and Xenopus Xlpou91 proteins (Figure 4.12a). This motif is very similar to a DEF (docking site for ERK, $^{F}/_{Y}$ -X-$^{F}/_{Y}$ -P) site sequence. Figure 4.12b shows examples of MAP kinase docking sites including the DEF site from human and yeast MAPK interacting proteins.

Figure 4.12c illustrates how a MAP kinase interacts with a docking site of one of its substrates. The MAPK docking sites are usually found near the target phosphorylation site on the substrate proteins, and a single docking site can enhance the phosphorylation of multiple target sites (Bardwell and Shah, 2006). As MAP kinases are proline directed kinases that phosphorylate the serine or the threonine in the dipeptide motif S/T-P (Bardwell and Shah, 2006), we looked for potential phosphorylation sites specific for MAP kinase that are present near the potential DEF site motif in the C-TD of Oct4, hOCT4, and Xlpou91. The existence of a potential MAP kinase phosphorylation site near the conserved motif present in the C-TD of Oct4, hOCT4, and Xlpou91 proteins, support the functionality of this motif as a DEF docking site for MAP kinases.

(a)

```
hOCT4          SSDYAQREDFEAAGSPFS-GGPVS--FPLAPGPHFGTPGYGS-PHFTALYSSVPFPEGEA 56
mOCT4          SIEYSQREEYEATGTPFP-GGAVS--FPLPPGPH-GTPGYGS-PHFTTLYS-VPFPEGEA 54
XlPou91        VYPYIRENGGEPYDTPQTLTPPSQGPFPLPQVMPSQVFPTVPLGANPTIYA-PTYHKNDM 59
XlPou60        -FRMSKGHEFVGGASPGS-IQSEHISFTPIPANSDYGLAS-LHPNRAPFYP-PPFPRNEL 56
XlPou25        GMPTVEENDGEGYDVAQTMGSPPVGHYALQQVVTPQGY-----MAAPQIYA-SAFHKNDL 54
DrPou2         ALPFDDECVEAQYYEQSPPPPPHMGGTVLPGGQGYPGPAH--PGGAPALYM-PSLHRPDV 57
AmOct4         -ICREEYDGFQQYPGMQP-GPPALSHLP--TSYIAQGYNG-AAAAFAAVYM-QPFHDSEM 54
                                .   .                      . .*   .    :
hOCT4          FPPVSVTTLGSPMHSN 72
mOCT4          FPSVPVTALGSPMHSN 70
XlPou91        FPQAMHHGIGMGNQGN 75
XlPou60        FPHMAP-GISMGVLTG 71
XlPou25        FPQTVPHGMAMGGHIG 70
DrPou2         FKNGLHPGL-VGHLTS 72
AmOct4         YSQTVSRHLHSN---- 66
               :       :
```

(b)

MAPK DOCKING SITES

D-site class:

		+++++ hXh
MEK1	MP	KKKPTP--IQL NPAPDG
MEK2	MLA	RRKPVLPALTI NPTIAE
MKK3	GKS	KRKKD---LKL SCMSKP
MKK6	SKG	KKRNPG--LKI PKEAFE
MKK4	MQG	KRKA----LKL NFANPP
MKK7-D1	REA	RRRID---LNL DISPQR
MKK7-D2	SPQ	RPRPT---LQL PLANDG
MKK7-D3	PPA	RPRHM---LGL PSTLFT
ySte7	TLQ	RRNLKG--LNL NLHPDV
yDig1	KSL	KRGRVPAPLNL SDSNTN
yDig2	HSL	KRKRVPPALNF SDIQAS
yFar1	MMS	KRGNIPKPLNL SKPISP
JIP1	DTY	RPKRPTT-LNL FPQVPR
JIP3	GRS	RKERPTS-LNV FPLADG
c-Jun	SNP	KILKQSMTLNL ADPVGS
ATF2	AVH	KHKHE---MTL KFGPAR
ELK1	QPQ	KGRKPRD-LEL PLSPSL
MKP1	TIV	RRRAKGA-MGL EHIVPN
MKP2	TIV	RRRAKGS-VSL EQILPA
MKP3	IML	RRLQKGN-LPV RALFTR
PTP-SL	LQE	RRGSNVS-LTL DMCTPG
HePTP	LQE	RRGSNVA-LML DVRSLG

MAPKAPK class:

		L +++ ++
MNK1	KSR	LARRRAL-- AQA
MNK2	QSK	LAQRRQR-- ASL
RSK1	SSI	LAQRRVRK- LPS
RSK2	RST	LAQRRGIKK ITS
yRck2	STI	LQRRKKV-- QEN

DEF-site class:

		FXFP
Elk1	AKLS	FQFP SSGS
c-fos	YTSS	FVFT YPEA
Fra1	FTPS	FTYP STPE
Fra2	SNLV	FTYP SVLE
c-Myc	PSVV	FPYP LNDS
Vinexin	RRSA	FPFP ITLQ
KSR	RDSR	FNFP AAYF
MKP1	TTTV	FNFP VSIP
yDig1	HIFA	FEFP LSSS

(c)

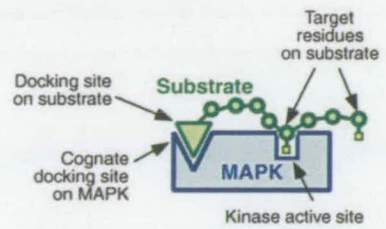


Figure 4.12- Carboxyl terminal alignment of PouV proteins reveals a potential DEF site in Xlpou91 , mouse and human Oct4

(a) Amino acid sequence alignments of the carboxyl terminal (C-TD) of PouV proteins uncover a potential sequence for the DEF site (MAP kinase docking site). The DEF site is coloured in blue. CLUSTAL W (1.83) multiple sequence alignment software was used for the alignment.
(b) MAPK doking sites from human and yeast MAPK-interacting proteins: D site, MAPKAPK, and DEF site class. (Bardwell and Shah., 2006)
(c) A schematic representing docking interaction between MAP kinase and one of its substrate. (Bardwell and Shah., 2006)

To check for the existence of any putative MAP kinase phosphorylation sites, the C-terminal domains of mouse Oct4, human OCT4, Xenopus Xlpou91, Xlpou25, and Zebrafish DrPou2 proteins were analysed for potential phosphorylation site using NetPhos.2.0 server. This software is based on neural network method, it predicts phosphorylation sites at serine (S), threonine (T) or tyrosine (Y) residues in eukaryotic proteins (Blom et al., 1999). The output of this analysis is shown for each C-TD individually in the following figures: mouse Oct4 C-TD (Figure 4.13a), human OCT4 C-TD (Figure.4.13b), Xenopus Xlpou91 C-TD (Figure 4.13c), Xlpou25 C-TD (Figure 4.13d), and Zebrafish DrPou2 (Figure 4.13e).

Taking into account the fact that MAP kinases phosphorylate the serine or the threonine in the dipeptide motif S/T-P, we identified MAP Kinase specific sites amongst all predicted phosphorylation sites.

This analysis shows that there are two putative MAPK phosphorylation sites in the C-TD of mouse Oct4 (T at position 15, and S at position 38), two sites in human OCT4 C-TD (S at position 15, and S at position 39), and one site in Xlpou91 C-TD (T at position 15), whereas, there was no MAP kinase specific site in C-TD of Xlpou25 or DrPou2. Figure 4.14 shows the positions of the potential MAP Kinase phosphorylation sites in the C-TD of our proteins in red.

(a)

```
   70 mOCT4
SIEYSQREEYEATGTPFPGGAVSFPLPPGPHGTPGYGSPHFTTLYSVPFPEGEAFPSVPVTALGSPMHSN
....S....Y....T.....................Y.S....T.........................
```

Phosphorylation sites predicted: Ser: 2 Thr: 2 Tyr: 2

Serine predictions

Name	Pos	Context	Score	Pred
		v		
mOCT4	1	----SIEYS	0.004	.
mOCT4	5	SIEYSQREE	0.955	*S*
mOCT4	23	GGAVSFPLP	0.015	.
mOCT4	38	PGYGSPHFT	0.974	*S*
mOCT4	46	TTLYSVPFP	0.073	.
mOCT4	57	EAFPSVPVT	0.042	.
mOCT4	65	TALGSPMHS	0.024	.
mOCT4	69	SPMHSN---	0.062	.
		^		

Threonine predictions

Name	Pos	Context	Score	Pred
		v		
mOCT4	13	EYEATGTPF	0.401	.
mOCT4	15	EATGTPFPG	0.940	*T*
mOCT4	33	GPHGTPGYG	0.496	.
mOCT4	42	SPHFTTLYS	0.032	.
mOCT4	43	PHFTTLYSV	0.758	*T*
mOCT4	61	SVPVTALGS	0.020	.
		^		

Tyrosine predictions

Name	Pos	Context	Score	Pred
		v		
mOCT4	4	-SIEYSQRE	0.215	.
mOCT4	10	QREEYEATG	0.944	*Y*
mOCT4	36	GTPGYGSPH	0.745	*Y*
mOCT4	45	FTTLYSVPF	0.160	.
		^		

(b)

```
   72 hOCT4
SSDYAQREDFEAAGSPFSGGPVSFPLAPGPHFGTPGYGSPHFTALYSSVPFPEGEAFPPVSVTTLGSPMHSN
...Y.........S.....................Y.S......Y.S.......................
```

Phosphorylation sites predicted: Ser: 3 Thr: 0 Tyr: 3

Serine predictions

Name	Pos	Context	Score	Pred
		v		
hOCT4	1	----SSDYA	0.002	.
hOCT4	2	---SSDYAQ	0.021	.
hOCT4	15	EAAGSPFSG	0.778	*S*
hOCT4	18	GSPFSGGPV	0.182	.
hOCT4	23	GGPVSFPLA	0.035	.
hOCT4	39	PGYGSPHFT	0.973	*S*
hOCT4	47	TALYSSVPF	0.019	.
hOCT4	48	ALYSSVPFP	0.613	*S*
hOCT4	61	FPPVSVTTL	0.059	.
hOCT4	67	TTLGSPMHS	0.030	.
hOCT4	71	SPMHSN---	0.062	.
		^		

Threonine predictions

Name	Pos	Context	Score	Pred
		v		
hOCT4	34	PHFGTPGYG	0.234	.
hOCT4	43	SPHFTALYS	0.017	.
hOCT4	63	PVSVTTLGS	0.030	.
hOCT4	64	VSVTTLGSP	0.471	.
		^		

Tyrosine predictions

Name	Pos	Context	Score	Pred
		v		
hOCT4	4	-SSDYAQRE	0.853	*Y*
hOCT4	37	GTPGYGSPH	0.745	*Y*
hOCT4	46	FTALYSSVP	0.695	*Y*
		^		

(c)

```
   74 XlPou91
VYPYIRENGGEPYDTPQTLTPPSQGPFPLPQVMPSQVFPTVPLGANPTIYAPTYHKNDMFPQAMHHGIGMGNQG
............Y.T..T....S.........................Y...Y...................
```

Phosphorylation sites predicted: Ser: 1 Thr: 2 Tyr: 3

Serine predictions

Name	Pos	Context	Score	Pred
		v		
XlPou91	23	LTPPSQGPF	0.890	*S*
XlPou91	35	QVMPSQVFP	0.223	.
		^		

Threonine predictions

Name	Pos	Context	Score	Pred
		v		
XlPou91	15	EPYDTPQTL	0.666	*T*
XlPou91	18	DTPQTLTPP	0.829	*T*
XlPou91	20	PQTLTPPSQ	0.310	.
XlPou91	40	QVFPTVPLG	0.010	.
XlPou91	48	GANPTIYAP	0.112	.
XlPou91	53	IYAPTYHKN	0.094	.
		^		

Tyrosine predictions

Name	Pos	Context	Score	Pred
		v		
XlPou91	2	---VYPYIR	0.086	.
XlPou91	4	-VYPYIREN	0.097	.
XlPou91	13	GGEPYDTPQ	0.723	*Y*
XlPou91	50	NPTIYAPTY	0.564	*Y*
XlPou91	54	YAPTYHKND	0.880	*Y*
		^		

(d)

```
   70 XlPou25
GMPTVEENDGEGYDVAQTMGSPPVGHYALQQVVTPQGYMAAPQIYASAFHKNDLFPQTVPHGMAMGGHIG
...T........Y.............................................................

Phosphorylation sites predicted:        Ser: 0  Thr: 1  Tyr: 1
```

Serine predictions

Name	Pos	Context	Score	Pred
		v		
XlPou25	21	QTMGSPPVG	0.234	.
XlPou25	47	QIYASAFHK	0.005	.
		^		

Threonine predictions

Name	Pos	Context	Score	Pred
		v		
XlPou25	4	-GMPTVEEN	0.516	*T*
XlPou25	18	DVAQTMGSP	0.461	.
XlPou25	34	QQVVTPQGY	0.087	.
XlPou25	58	LFPQTVPHG	0.078	.
		^		

Tyrosine predictions

Name	Pos	Context	Score	Pred
		v		
XlPou25	13	DGEGYDVAQ	0.860	*Y*
XlPou25	27	PVGHYALQQ	0.043	.
XlPou25	38	TPQGYMAAP	0.325	.
XlPou25	45	APQIYASAF	0.178	.
		^		

(e)

```
  72 DrPou2
ALPFDDECVEAQYYEQSPPPPPHMGGTVLPGGQGYPGPAHPGGAPALYMPSLHRPDVPKNGLHPGLVGHLTS
............Y............................................................

Phosphorylation sites predicted:        Ser: 0  Thr: 0  Tyr: 1
```

Serine predictions

Name	Pos	Context	Score	Pred
		v		
DrPou2	17	YYEQSPPPP	0.019	.
DrPou2	51	LYMPSLHRP	0.031	.
DrPou2	72	GHLTS----	0.006	.
		^		

Threonine predictions

Name	Pos	Context	Score	Pred
		v		
DrPou2	27	HMGGTVLPG	0.248	.
DrPou2	71	VGHLTS---	0.302	.
		^		

Tyrosine predictions

Name	Pos	Context	Score	Pred
		v		
DrPou2	13	VEAQYYEQS	0.174	.
DrPou2	14	EAQYYEQSP	0.928	*Y*
DrPou2	35	GGQGYPGPA	0.111	.
DrPou2	48	APALYMPSL	0.045	.
		^		

Figure 4.13- Identification of potential phosphorylation sites in the C-TD of PouV proteins

C-TDs of (a) mouse Oct4, (b) human Oct4, (c) Xlpou91 (d) Xlpou25, and (e) DrPou2 proteins were analysed for potential phosphorylation site using NetPhos.2.0 server.

The software produces predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins.

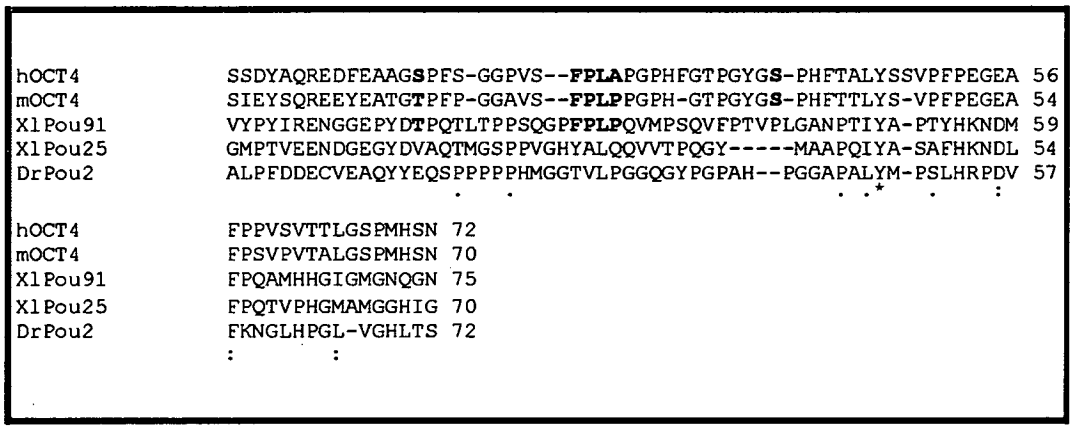


Figure 4.14-Positions of potential MAPK phosphorylation sites in the C-TD of PouV proteins

This analysis shows that there are two putative MAPK phosphorylation sites in the C-TD of mouse Oct4 (T at position 15, and S at position 38), two sites in human Oct4 C-TD (S at position 15, and S at position 39), and one site in Xlpou91 C-TD (T at position 15), whereas, there was no MAP kinase specific site in C-TD of Xlpou25 or DrPou2 proteins.
Amino acid sequence alignments of the carboxyl terminal (C-TD) of PouV proteins uncover a potential sequence for the DEF site (MAP kinase docking site). The DEF site is coloured in blue. CLUSTAL W (1.83) multiple sequence alignment software was used for the alignment.

As these sites are predicted MAP kinase phosphorylation targets, we tested their relevance to PouV activity by mutating the putative MAP kinase docking site in both Xlpou91 and mouse Oct4. The critical conserved residues of the DEF docking site were mutated from FPLP to AAAP (Figure 4.15a1). The resulting mutant proteins were tested for their transcriptional activity and their ability to rescue *Oct4* null ES cell phenotypes. Table 4.3 summarises the ability of these mutants proteins to activate transcriptions and to rescue ES cell self-renewal.

Both mutant Oct4 and Xlpou91 proteins appear to have identical activity to wild type proteins when tested for their ability to activate transcription from both the octamer binding motif luciferase reporter (Figure 4.15b1), and the *Fgf4* enhancer luciferase reporter (Figure 4.15b2), and (Figure 4.15b3). Similarly, when tested for their ability to rescue ES cell self-renewal, the mutant proteins appear able to support the growth of normal undifferentiated ES cell colonies (Figure 4.16b1 and b2). However, when the quantitative ability of these mutant proteins to rescue self-renewal is measured, some reduction in their rescue indices is observed (Figure 4.16a). Suggesting that while the MAPK docking domain appeared non-essential for the support of ES cell self-renewal, it is a likely contributing factor to the ability of the C-TD to increase the efficiency of quantitative rescue when transferred from Xlpou91 to either Xlpou25 or DrPou2.

4.2.7. Investigating the importance of Xlpou91 POU domain

4.2.7.1. Alignment of DNA binding POU domains of PouV proteins reveals possible conserved residues responsible for maintaining the undifferentiated phenotype of ES cells

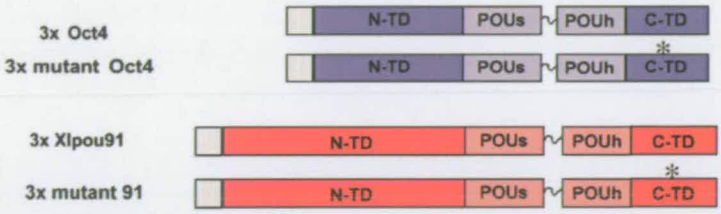
Xlpou91 POU domain appears essential for maintaining the undifferentiated phenotype of ES cells. This is in accordance with previous findings showing that the POU domain of Oct4 is responsible for conferring its unique function in ES cells (Niwa, Masui et al. 2002; Nishimoto, Miyagi et al. 2005).

(a.1)

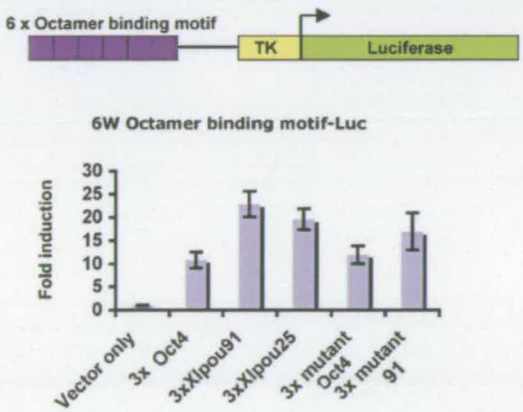
```
mOCT4          SIEYSQREEYEATGTPFP-GGAVS--AAAPPGPH-GTPGYGS-PHFTTLYS-VPFPEGEA 54
Xl Pou91       VYPYIRENGGEPYDTPQTLTPPSQGPAAAPQVMPSQVFPTVPLGANPTIYA-PTYHKNDM 59

mOCT4          FPSVPVTALGSPMHSN 70
Xl Pou91       FPQAMHHGIGMGNQGN 75
```

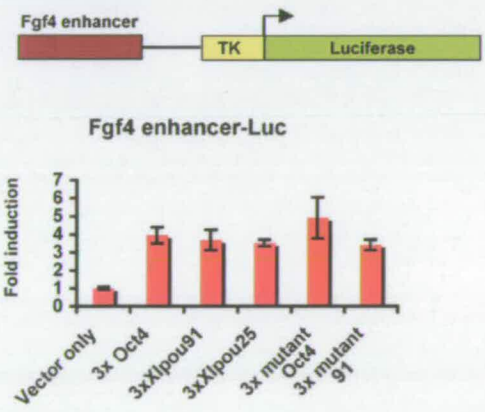
(a.2)



(b.1)



(b.2)



(b.3)

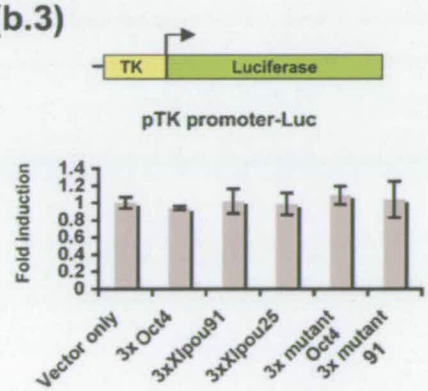


Figure 4.15-Cloning and transcriptional activity of Oct4 and Xlpou91 proteins with mutated DEF site

(a.1) Mutating the DEF site sequence from FPLP to AAAP sequence. (a.2) Schematic representation of Flag tagged wild type and mutated Oct4 and Xlpou91 proteins.

(b) 3xFlag- mutated Oct4 and Xlpou91 proteins activate transcription in a specific manner from two luciferase reporters: one contains multiple copies of the octamer binding motif **(b.1)** and the other contains the Fgf4 enhancer **(b.2).** A luciferase reporter containing the thymidine kinase (TK) promoter was used to test the specificity of the proteins transcriptional activity **(b.3).**

The indicated cDNAs were cotransfected with luciferase reporters. Fold induction represents the increase in transcription compared with the empty vector control. Data represent the mean value of two independent experiments.

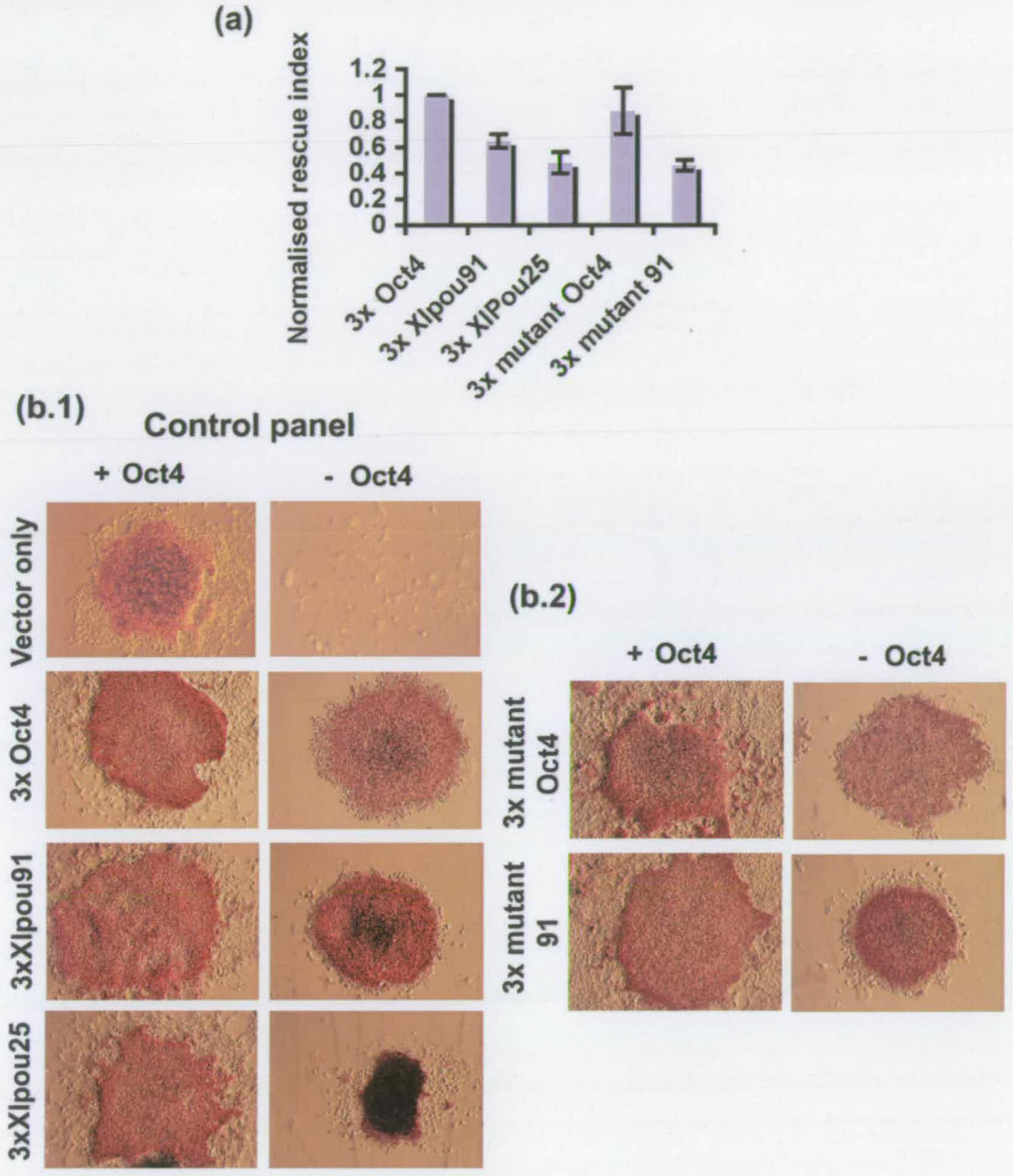


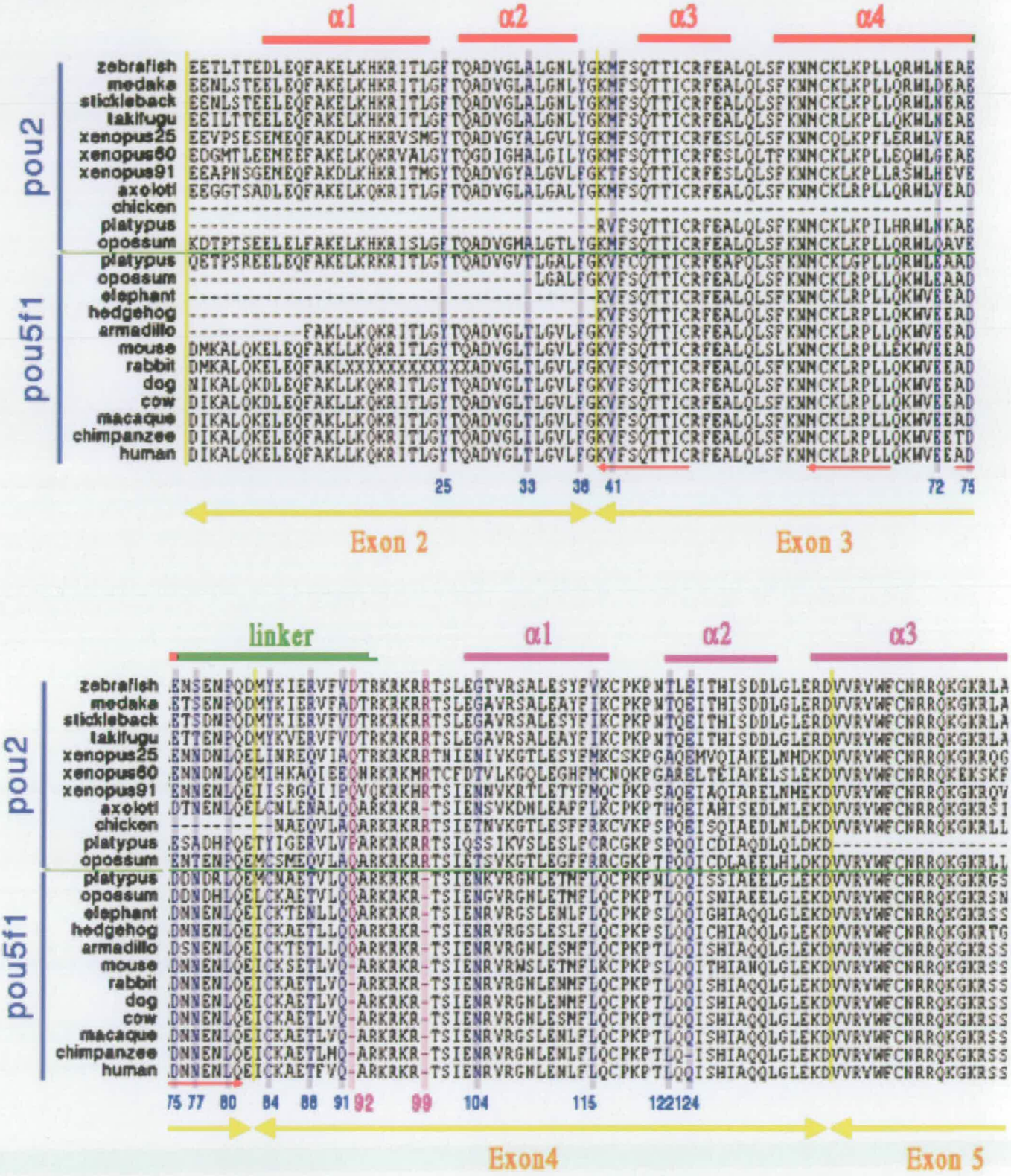
Figure 4.16-Mutating the DEF site has little effect on ES cells self-renewal rescue ability of Oct4 and Xlpou91 proteins

(a) Rescue index for 3x Flag Oct4 and Xlpou91 proteins with mutated DEF site. The rescue index values were normalised to 3xFlag Oct4 value. Data represents the mean values obtained from two independent experiments. **(b)** Morphology of colonies rescued by the mutated proteins. Morphology of representative colonies rescued by wild type PouV proteins **(b.1)**, and proteins with mutated DEF site **(b.2)** is shown.

These studies showed that a chimeric protein consisting of the activation domain of Oct6 and the POU domain of Oct4 could rescue the self-renewal of *Oct4* null ES cells. They reported that the threonine residue at position 22 in the α1-helix of the POU specific domain and the linker region of Oct4 are the critical elements responsible for the function of the POU domain in ES cells (Nishimoto, Miyagi et al. 2005). However, they did not address the consequence of mutating these elements in the context of Oct4.

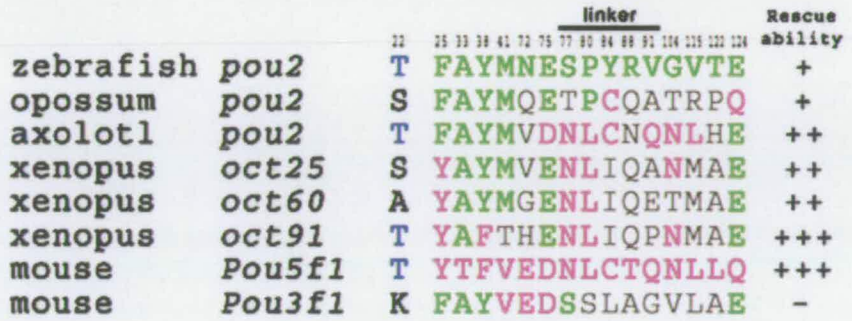
In order to verify the importance of the elements identified in the study by Nishimoto and colleagues, and further uncover other elements in the linker region that are critical for ES cell self-renewal, we aligned the amino acid sequences of POU domains from various PouV proteins differing in their ability to maintain ES cell self-renewal. We aimed to identify residues that conserved only in PouV proteins with the ability to maintain the undifferentiated phenotype of ES cells. During the course of this study, Niwa and colleagues published an analysis of Oct4 evolution (Niwa et al., 2008), which included an investigation of functional differences between Oct4 and more primitive PouV proteins, and focused on key conserved positions that vary in PouV protein evolution. Their analysis, represented in Figure 4.17a, and b highlights the differences in amino acid sequences of POU domains of Zebrafish pou2, Opossum pou2, Axolot Oct4, Xenopus Xlpou25, Xlpou60, Xlpou91, Platypus Oct4, mouse Oct4 and Oct6 proteins. Using this alignment, Niwa et al 2008 predicted the following amino acids T22, Y25, T33, F38, V41, and T88 as potential mediators of Oct4 specificity in ES cells. The importance of these residues for Oct4 function was evaluated by mutating them one amino acid at the time in Oct4 to match their corresponding residues found in Pou2 homologues, and testing the abilities of the resulting Oct4 variants to rescue ES cell self-renewal in the complementation assay (Niwa et al., 2008). Interestingly, none of these variants had a significant reduction in their ability to rescue ES cell self-renewal in comparison to wild type Oct4. As a result, Niwa et al concluded that individual amino acid changes are not sufficient to eradicate the ES cell specific activity of Oct4 (Niwa et al., 2008).

(a)



(Niwa et al., 2008)

(b)



(Niwa et al., 2008)

Figure 4.17-Alignment of the amino acid sequences of the POU domains of the vertebrate PouV proteins

(a) and **(b)** are figures from the paper published by Niwa et al 2008.

Figure **(a)** represents alignment of the amino acid sequences of POU domains of vertebrate PouV proteins. As determined by the authors of the paper the upper red lines define the α-helix structures of the POUs and POUh domains, and the green line shows the linker domain. The lower yellow horizontal arrows and vertical lines show the lenght and boundaries of each exon.
The blue vertical lines indicate amino acid residues different between Pou2 and Pou5f1 orthologues, and the red vertical lines show deletions in either some or all Pou5f1 orthologues.
Figure **(b)** shows residues that are different between POU domains of Pou2 homologues and Pou5f1 orthologues and the rescue ability of each protein. The numbers on the top indicate the position of each residue in the alignment shown in figure **(a).** Green letters show residues identical to those of the Zebrafish pou2 protein, and red letters show residues identical to those of mouse Pou5f1 protein. Blue letters indicate the threonine residue which was identified by Nishimoto as important for Pou3f1 function (Niwa et al, 2008).

While a single amino acid might not be able to eliminate Oct4 function in ES cells we have identified a single amino acid change in some DrPou2 sequences that eradicates completely the ability of this protein to rescue *Oct4* null ES cell self-renewal. As stated in the beginning of this chapter, we have identified two EST sequences for the *DrPou2* gene. The original *DrPou2* sequence used by our group has an alanine (DrPou2(A)) in place of a threonine (DrPou2(T)) in the POU domain at position 348 in the protein (Figure 4.18), which corresponds to position 228 in the Oct4 sequence. In fact this T to A substitution in the DrPou(A) sequence may represent a polymorphism of *DrPou* gene that is present in some populations of the Zebrafish specie or could be simply a result of a cloning artefact. Further investigation is required to determine the nature of this amino acid substitution in the protein.

In the proceeding sections I have presented data based on the DrPou2(T), that has a self-renewal rescue ability similar to that of Xlpou25 protein. However, in this section I compare these results to those obtained with DrPou2(A) that has no capacity to rescue Oct4 activity in ES cells.

The ability of DrPou2(T) to substitute for Oct4 activity in ES cells, albeit at a very reduced level, suggests that this threonine residue at position 348 is critical, and mutating it is sufficient to abolish the function of DrPou2 protein in ES cells. To investigate this possibility, and to ask whether Xlpou91 C-TD would have any positive effect on the rescue ability of DrPou2(A) protein, we generated a series of chimeric proteins between DrPou2(A) and Xlpou91 by swapping the domains of the two proteins as explained in previous sections (Figure 4.19a). Table 4.4 summarises the ability of these chimeric proteins to activate transcription and rescue ES cell self-renewal.

These Xlpou91-DrPou2(A) chimeric proteins were tested for their abilities to activate transcription from luciferase reporters harbouring the *Fgf4* enhancer or six copies of the octamer-binding motif, and were all found to specifically activate transcription from both reporters in an identical fashion to their Xlpou91-DrPou2(T) counterparts (Figure 4.19b1, b2 and b3).

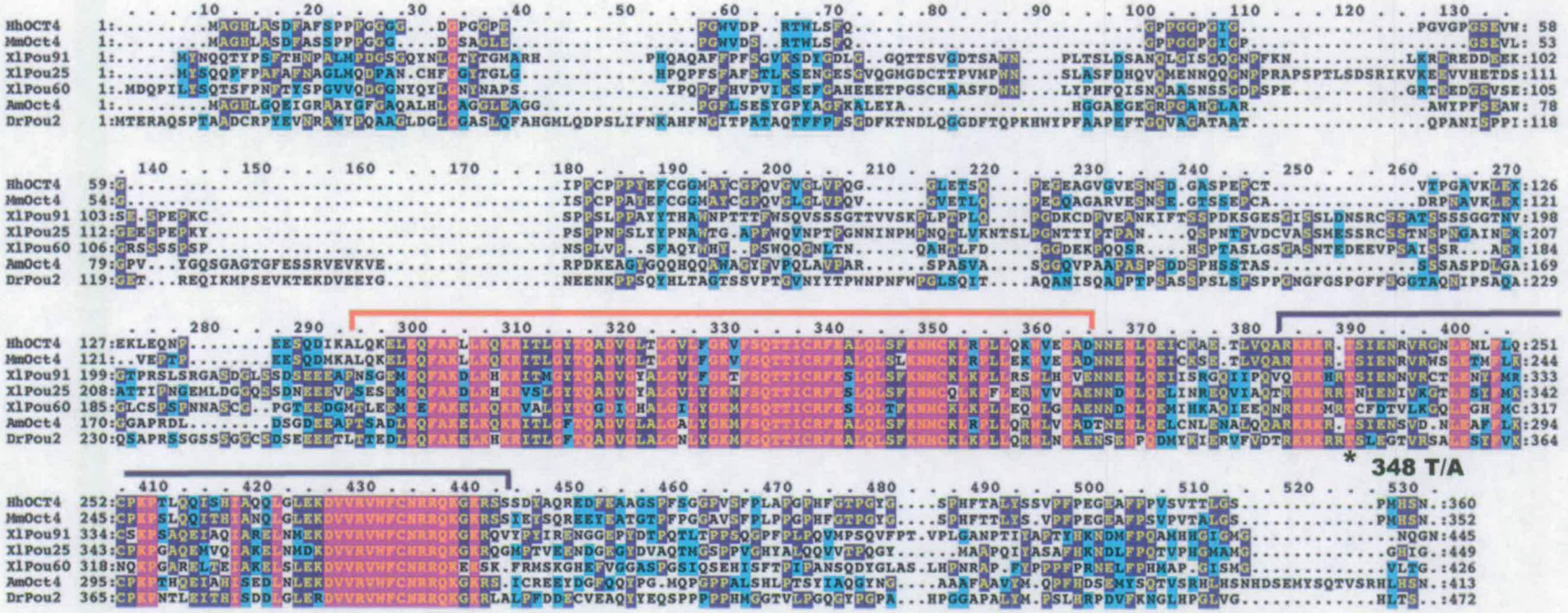


Figure 4.18 - The difference between DrPou2 (T) and DrPou2(A) proteins

The zebra fish DrPou2(A) protein has an Alanine instead of the Threonine at position 348 of the amino acid sequence of the protein. This threonine residue is conserved in human, mouse, Xenopus and axolotl PouV proteins.
PouV proteins alignment as published in Morrison and Brickman., 2006. Alignment of human (hOCT4), mouse (mOct4), Xenopus (Xlpou91, Xlpou25, Xlpou60), axolotl (AmOct4) and zebrafish (DrPou2) PouV proteins. (Alignment from Morrison and Brickman., 2006). POU specific (POUs) and POU homeodomain (POUh) domains are boxed in red and blue respectively. Pink indicates identical residues, light blue indicates similar residues, and dark blue indicates conservation within a subset of residues.

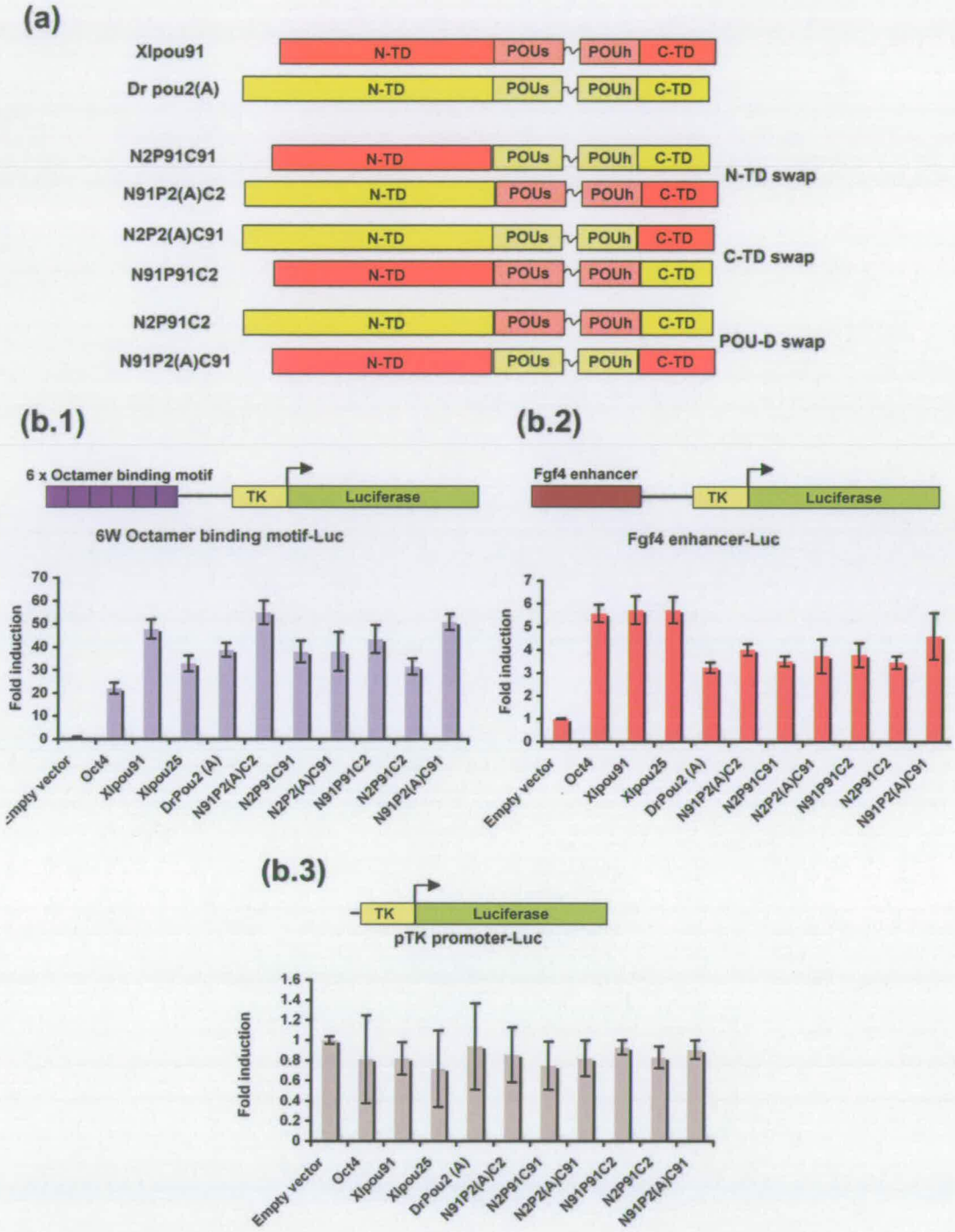


Figure 4.19–Cloning and transcriptional activity of Xlpou91-DrPou2(A) Chimeric proteins

(a) Schematic representation of Xlpou91-DrPou2(A) chimeric proteins: chimeric proteins were generated by swapping either the amino acid terminal (N-TD), the carboxyl terminal (C-TD) or the POU domain of Xlpou91 with DrPou2(A) domains. **(b)** Xlpou91-DrPou2(A) chimeric proteins are able to activate transcription in a specific manner from two luciferase reporters: one containing multiple copies of the octamer binding motif **(b.1)** and the other one containing the Fgf4 enhancer **(b.2).** A luciferase reporter containing the thymidine kinase (TK) promoter was used to test the specificity of the chimeric proteins transcriptional activity **(b.3).**

The indicated cDNAs were cotransfected with luciferase reporters. Fold induction represents the increase in transcription compared with the empty vector control. Data represent the mean value of two independent experiments.

When these Xlpou91-DrPou2(A) chimeric proteins were tested for their ability to rescue ES cell self-renewal, all proteins harbouring the POU domain from the DrPou2(A) protein (DrPou2(A), N91P2(A)C2, N2P2(A)C91, N91P2(A)C91) were unable to support undifferentiated ES cell growth, and therefore showed no rescue ability (Figure 4.20a, b1, b2, b3, and b4).

These results reaffirm that this threonine in the POU specific domain of DrPou2 protein is essential for any ES cell activity, and that without a POU domain, the C-TD is not able to confer any advantageous effect on the rescue ability of the protein.

Table 4.4- Summary of the ability of the different Xlpou91 and DrPou2(A) chimaeric proteins to activate transcription and to rescue ES cell self-renewal.

The transcriptional activity of the different Xlpou91-DrPou2(A) chimaeric proteins is assessed based on their ability to activate transcription from luciferase reporters harbouring Oct4 dependent sequences either from the Fgf4 enhancer or 6 reiterated copies of the Octamer binding motif. The ability of the different Xlpou91-DrPou2(A) chimaeric proteins to rescue ES cell self–renewal is measured based on the morphology of the rescued colonies and the number of the rescued colonies represented as the rescue index. N: N-terminus, C:C-terminus, P:POU domain, 91: Xlpou91, 2: DrPou2(A). Rescue index value is the mean from two independent experiments.

	Chimeric proteins	Transcriptional activity		Morphology of rescued colonies	Rescue index
		6W Octamer binding motif	Fgf4 enhancer		
Wild type proteins	Oct4	+ +	+ + +	Typical undifferentiated ES cell colonies	1
	Xlpou91	+ + + +	+ + +	Typical undifferentiated ES cell colonies	1.02
	DrPou2(A)	+ + +	+ +	Highly differentiated colonies	0.02
N terminal swap	N91P2(A)C2	+ + +	+ +	Highly differentiated colonies similar to DrPou2(A) expressing colonies	0.07
	N2P91C91	+ + +	+ +	Undifferentiated colonies similar to Xlpou91 rescued colonies, with slight increase in differentiated cells around the periphery	1.85
C terminal swap	N2P2(A)C91	+ + +	+ +	Highly differentiated colonies similar to DrPou2(A) expressing colonies	0.07
	N91P91C2	+ + +	+ +	Undifferentiated ES cell colonies that are slightly smaller than Xlpou91 rescued colonies	0.52
POU domain swap	N2P91C2	+ + +	+ +	Undifferentiated colonies similar to Xlpou91 rescued colonies, with slight increase in differentiated cells around the periphery	0.95
	N91P2(A)C91	+ + +	+ +	Highly differentiated colonies similar to DrPou2(A) expressing colonies	0.06

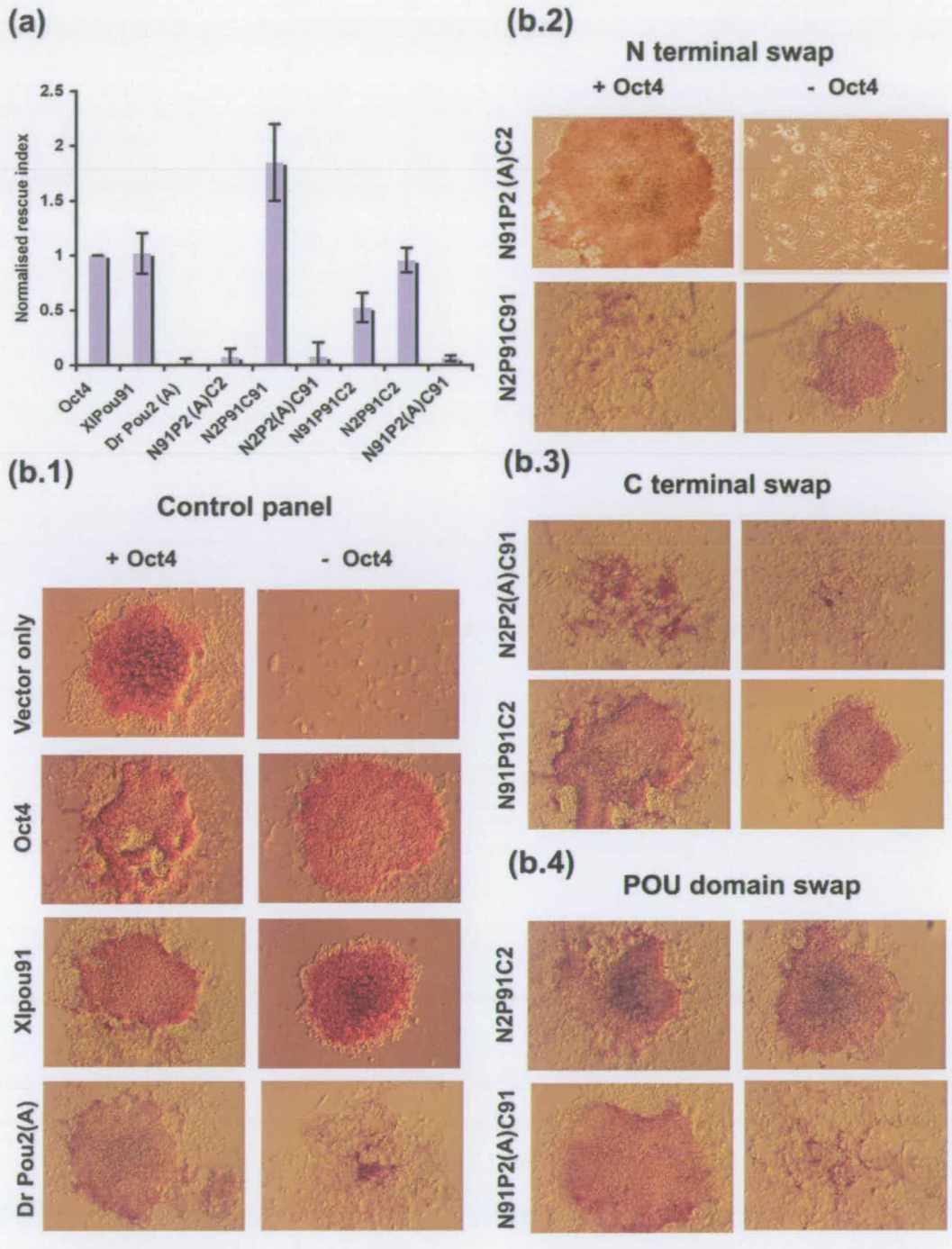


Figure 4.20- The inability of Zebrafish DrPou2(A) protein to rescue ES cells self-renewal reveals the crucial role of POU domain of the PouV proteins in the process

(a) Rescue index for Xlpou91-DrPou2(A) chimeric proteins. Rescue index values were normalised to Oct4 value. Data represents the mean values obtained from two independent experiments. **(b)** Morphology of colonies rescued by the Xlpou91-DrPou2(A) chimeric proteins: morphology of representative colonies rescued by wild type PouV proteins **(b.1)**, chimeric proteins generated by N terminal swap **(b.2)**, C terminal swap **(b.3)**, and POU domain swap **(b.4)**.

4.3. Discussion

The mechanism by which pluripotency is established and maintained is still not fully understood. Oct4 was identified as a crucial factor governing pluripotency and self-renewal in both early development and ES cells. Although Oct4 function has been studied extensively, the mechanism by which it controls pluripotency and self-renewal is not clear yet. My studies have shown that the remarkable ability of Xlpou91 to rescue ES cell self-renewal is encoded by two functions. An activation domain in the C-TD that is required for the efficiency of ES cell self-renewal (although this activity has become somewhat redundant in Oct4), and crucial sequences within the POU domain that are essential for the maintenance of the undifferentiated phenotype of ES cells.

In fact, previous studies tried to investigate the functional importance of each domain of the Oct4 protein. The activation domains of Oct4 have been characterized via fusion to heterologous DNA binding domains. Imagawa et al 1991 fused the N-TD or C-TD Oct4 to the DNA binding domain of c-Jun and tested their abilities to activate transcription from reporter genes containing reiterated c-Jun binding sites in P19 EC cells. Based on this analysis, only the N-TD was believed to have activation function (Imagawa et al., 1991). However, when the activation function of Oct4 was assayed in a native context from reiterated octamer sites, Vigano et al 1996 reported that both the N and C terminal domains are able to activate transcription in Hela cells. They also reported that the C-TD was the stronger activation domain (Vigano and Staudt 1996).

Brehem et al 1997 confirmed the trans-activation ability of both the N and C-TD and reported that the activity of the C domain is cell type specific and is mediated by the POU domain. In fact, when they compared the activity of these domains when fused to either Gal4 or Pit1 DNA binding domains, they found that Gal4-C-TD fusions activated transcription from a reporter containing 6 copies of the octamer binding motif in Hela cells, but failed to activate transcription in 923 or NIH3T3 cell lines. However, a Pit1-C-TD fusion had the capacity to activate transcription of this

reporter regardless of the cell lines used (Brehm, Ohbo et al. 1997). These findings are consistent with the requirement for Sox2 at the *Fgf4* promoter to enable the C-TD to activated transcription (Ambrosetti et al., 2000).

Interestingly, while this work is all specific to transcriptional activation in reporter assays, it does point to a specific role for the C-TD in the regulation of transcription by POU proteins, which in fact is in accordance with my observations about its contribution to the Oct4-like function in ES cells.

Although these studies provided valuable insight into the function of the different domains of Oct4 proteins, they provide no insight into the physiological roles of these domains in establishing and maintaining pluripotency and self-renewal.

In 2002, Niwa and colleagues addressed this question and tried to map the functional domains of Oct4 that are necessary for maintaining pluripotency and self-renewal in ES cells. They found that Oct4 proteins lacking either the N or C-TD were able to rescue ES cell self-renewal in their complementation assay. They also show that fusing the POU domain of Oct4 to a heterologous trans-activation domain is sufficient for maintaining ES cell self-renewal, and therefore concluded that the POU domain of Oct4 is uniquely required and coupled with a generic proline rich trans-activation domain is sufficient to maintain the undifferentiated stem cell phenotype (Niwa et al., 2002). However, this argument for generic activity is difficult to reconcile with their findings that the activity of N and C trans-activation domains of Oct4 regulate different sets of genes.

My data clearly shows that the C-TD is not required in Oct4, but is absolutely required for Xlpou91, suggesting that Oct4 might have redundant functions that mask the specific role for the C-TD. Moreover, the original studies by Niwa et al 2002 implied that there was a more important role for the C-TD than the N-TD in ES cell self-renewal. They reported that transfecting *Oct4* null ES cells with an Oct4 protein lacking the C-TD produced only the third of ES cell colony number obtained with an Oct4 protein missing the N-TD only (Niwa et al., 2002), indicating that even in the context of Oct4, this domain is important.

The possibility that the C-TD is a stronger activation domain than the N-TD has been reported previously (Vigano and Staudt, 1996) and could explain why the C-TD appears to have a more specific role. However, I have never observed a significant correlation between transcriptional activation and rescue of *Oct4* null ES cell phenotype. Instead, I favor a role for MAP kinase in the regulation of Oct4 activity as it has already been implicated in ES cell self-renewal (Nichols et al., 2009). While the phenotypes I observed were not strong, I have not yet shown that I have effectively neutralized the MAPK phosphorylation of the C-TD in the docking site mutants. In addition to that, MAPK phosphorylation in unlikely to be the only factor regulating C-TD activity as Xlpou25 and DrPou2(T) have some residual activity in their C-TD despite the absence of the consensus MAPK docking site in their C-TD sequences.

In addition to the role of the C-TD, we also showed that there is a potential negative role for the Xlpou91 N-TD in regulating Oct4-like activity. It appeared that chimeras without the Xlpou91 N-TD showed higher rescue indices than the full length Xlpou91. Interestingly, colonies rescued by these chimeric proteins had a slightly higher level of background differentiation. Niwa et al 2002 reported that the N terminal domain of Oct4 protein is able to direct the expression of *Eb'af/Lefty1* gene and that the C terminal domain is specifically lacks this ability (Niwa et al., 2002). As Lefty is a negative regulator of Nodal signalling, one possible mechanism by which the N-terminus could be dampening the rate of self-renewal as reflected in the rescue index might be through Lefty mediated Nodal antagonism.
Although the effect of Xlpou91 N-TD on the rate of ES cell self-renewal is very interesting, it was only detected in swap experiments and straight deletion of the N-TD resulted in a reduction of self-renewal efficiency in both Xlpou91 and Oct4 itself.

The major determinant of PouV proteins ability to support ES cell morphology and undifferentiated phenotype appears to be the POU domain. Studies focusing on the difference between Class III and Class V POU proteins indicated that the crucial region of PouV protein responsible for supporting ES cell growth is the first α-helix

of the POU specific domain and the linker region joining the POU specific and the POU homeodomain. They identified the threonine residue at position 22 in the α1-helix of POU domain as the critical element responsible for ES cell self-renewal in this region. Although they did not show that any specific residue was essential for Oct4 activity (Nishimoto et al., 2005; Niwa et al., 2002).

In our laboratory, we took a slightly different approach to address this question. We based our investigation on previous findings from our group. These findings indicate that Oct4 function in maintaining pluripotency and self-renewal in mammals is derived from a conserved role of PouV proteins to prevent premature commitment of cells in the developing embryo, and that only some of these PouV proteins are able to support ES cell self-renewal in the absence of Oct4 (Morrison and Brickman, 2006).

In This study, we exploited the sequence similarity and the functional differences between the very closely related Xenopus proteins Xlpou91 and Xlpou25 to identify functional domains responsible for maintaining the undifferentiated phenotype of ES cells. The comparison of very closely related proteins rather than proteins from different subclasses facilitates the identification of the functional domains because their sequences are more conserved.

Our study showed that even within the class V family of POU protein, the POU domain is the major determinant required for the support of ES cell undifferentiated morphology. Moreover, this study also shows that, despite their functional differences in ES cells, these PouV proteins all appear to activate transcription from reporters bearing Oct4 responsive elements to the same degree. Which in fact suggest that functional differences between these PouV proteins in ES cells are more likely to be caused by their different abilities to form promoter specific multimeric protein-protein complexes, rather than their ability to recognize target DNA sequences. In fact, the POU domain was shown to affect Oct4 function in many aspects. In addition to DNA binding it has been shown to be involved in protein-protein interactions, dimerization ability, and stability. The POU domain consists of two DNA binding domains connected with a flexible linker. This structure allows the protein to form heterodimers with other transcription factors and homodimers in

different conformations (PORE or MORE confirmation) depending on the configuration of the target octamer binding site, and therefore enable it to interact with a variety of target sequences (Botquin, Hess et al. 1998; Remenyi, Lins et al. 2003). The POU domain of Oct4 protein is also involved in protein-protein interactions and the selective recruitments of co-factors. It mediates the interaction with Ets2, adenovirus E1A, and Sox2 proteins (Ezashi et al., 2001; Scholer et al., 1991; Yuan et al., 1995), and was shown to be vital for the synergistic activation of Fgf4 enhancer by the Sox2-Oct4 complex (Ambrosetti et al., 2000). Furthermore, different dimmer conformations lead to different recruitment of transcriptional co-activators. It was shown that OBF-1 binds to the PORE configuration of the Oct1 dimer and synergises with it in transcriptional activation, but fails to bind Oct1 in the MORE conformation (Tomilin, Remenyi et al. 2000).

In addition to protein-protein interactions, the presence of specific sequences within the POU domain might effect the post-translational modification of the protein. Recent studies reported that Oct4 function could be regulated by post-translational modifications either within or mediated by the POU domain of the protein. Zhang et al 2007 showed that Oct4 interacts with Ubc9 and E2 conjugation enzymes to mediate sumoylation of lysine 118 through its POU domain, and that this modification is required for optimal levels of self-renewal in ES cells (Zhang et al., 2007). Moreover, Saxe and colleagues (2009) reported that phosphorylation of Oct4 at serine 229, which is adjacent to the POU homeodomain, partially controls its transactivation activity (Saxe et al., 2009). Interestingly, phosphorylation of Ser229 prevents DNA binding of Oct4 on PORE sequences, but does not alter its binding ability to other octamer motif configurations (Saxe et al., 2009).

Because the function of Oct4 is affected by its POU domain in many ways, further investigations of the different POU domain sequences described here will be necessary to determine the molecular mechanism by which they regulate PouV protein specificity in ES cells.

During the course of our study, Niwa and colleagues published their study utilizing a

similar strategy to ours (see result section for more details). They compared amino acid sequence of POU domains from Oct4 homologues. They reported that mutating single amino acids did not have a significant effect on Oct4 ability to rescue *Oct4* null ES cells self-renewal. Therefore, they concluded that changing individual amino acid residues is not sufficient to abolish the ability of Oct4 to maintain ES cell self-renewal, and suggested that the combination of all these amino acids is required for its function (Niwa et al., 2008). However, we found that changing the Threonine residue at position 348 in DrPou2 protein with an Alanine resulted in eliminating the ability of the protein to rescue ES cell self-renewal entirely.

Investigating this finding might provide a better understanding of Oct4 regulated molecular mechanism underlying the maintenance of ES cells, especially that this one amino acid change did not cause a loss of general DNA binding ability of the protein. However, before embarking on examining the effect of this change on the specific protein/protein, protein/DNA interaction, dimerization or other possible post-translational modifications of the protein, it is necessary to confirm the effect of substituting this threonine residue in the other PouV proteins especially Oct4 and Xlpou91.

CHAPTER 5

General Discussion

5. General Discussion

5.1. Oct4 maintains self-renewal through its activator function

One of the major findings of this study is that the activator form of Oct4, but not the repressor form, is able to maintain ES cells in an undifferentiated state. This finding is in disagreement with earlier models in which Oct4 maintains pluripotency by acting mainly as a direct blocker of differentiation (Pan et al., 2002). Our findings do not refute the key importance of inhibiting differentiation by Oct4, but point to a direct and primary role for Oct4 in activating pluripotency genes.

In this study, we engineered Oct4 proteins that acts as activators (Oct4λVP2) or repressors only (Oct4λEnR), and tested their abilities to maintain uncommitted cells both *in vivo* and *in vitro*. The repressor form of Oct4 was not able to maintain ES self-renewal, and induced differentiation in both wild type ES cells and Xenopus embryos. In fact, the expression of the repressor form of Oct4 in Xenopus embryos induced the expression of differentiation-associated genes *Gsc* and *Mixer,* and repressed gene expression associated with uncommitted cells (e.g. *Bmp4*). However, the activator-only Oct4 (Oct4λVP2) both maintained self-renewal in ES cells, and inhibited differentiation genes in Xenopus embryos (e.g. *Gsc* and *Mixer*). This data pointed to the importance of the activator function of Oct4 in maintaining uncommitted cell populations both *in vivo* and *in vitro*.

While it is still possible that Oct4λVP2 can repress certain genes in ES cells, we tested its function on a number of reporter genes and found that under no conditions did it function as a repressor. Moreover, our microarray data revealed that in Oct4λVP2 cell line, a proportion of genes that are normally believed to be repressed by the Nanog and Oct4 complex in ES cells (Loh et al., 2006) are in fact upregulated, indicating that Oct4λVP2 is a potent activator and overrides Nanog repression of these genes despite the later high expression levels in this cell line.

In the light of this finding, a closer analysis of earlier studies shows that gene activation of pluripotency genes is the primary mechanism for maintaining ES cell self-renewal. Activated Oct4 targets play an integral part in the core transcriptional circuitry of pluripotency (Zhou et al., 2007) and our data suggests that the activity of this network is dominant over the inadvertent expression of lineage markers.

In this study we provide the first evidence, to our knowledge, that activation by Oct4 is necessary and sufficient for maintaining self-renewal. A rigorous demonstration that this was sufficient for pluripotency would require the generation of a cell line in which the endogenous *Oct4* had been replaced by *Oct4λVP2* and the demonstration that these demonstrate germ line transmission.

5.2. Oct4 maintains pluripotency by upregulating pluripotency genes directly and inhibiting differentiation genes indirectly

In Xenopus embryo, *Bmp4* seems to be a potential mediator of PouV protein function to suppress differentiation. *Bmp4* expression seems dependent on PouV expression, injection of *Oct4λVP2* mRNA elevated *Bmp4* levels, whereas, Oct4λEnR overexpression decreased its expression levels. Moreover, injection of *Bmp4* mRNA was able to rescue *PouV* knockdown phenotype.

Similarly, we found that in ES cells Oct4λVP2 levels correlated with *Nanog* expression levels. *Nanog* was the only core pluripotency gene that appeared to respond to Oct4λVP2 levels. Interestingly, while no Nanog homologue has been identified in Xenopus, its closest relative would probably be the Nkx family member Xvent1, a BMP target that depends on PouV proteins for its expression (Cao et al., 2004).

Oct4 has been recently shown to regulate an extended *Nanog* locus (Levasseur et al., 2008), in which reside *Dppa3* and *Gdf3*. While we have not shown direct binding of Oct42 λVP2 to this region, we did find that the over-expressed genes in Oct4 λVP2 expressing cells map to specific genomic sites and that this locus represents one of

these clusters. In fact, these clusters of upregulated genes may represent key regulatory nodes in the pluripotency network.

Our results also show that Oct4λVP2 represses key developmental genes such as *Cdx2* and other putative Oct4 repressed targets (Loh et al., 2006). This supports the notion that the upregulation of these genes in response to Oct4 knockdown is probably an indirect consequence of differentiation. While it would seem unlikely that *Cdx2* is directly repressed by Oct4λVP2, it is directly repressed by Nanog (Chen et al., 2009). Nanog has also been shown to repress other differentiation genes such as *Eomes, Otx2* and *Msx2* in genome-wide ChIP analysis (Chen et al., 2009; Kim et al., 2008). It is therefore likely that Nanog directly represses the lineage-specific genes downstream of Oct4λVP2.

In summary, we think that Oct4λVP2 activate pluripotency regulators such as Nanog, which in turn will lead to the repression of differentiation specific genes. We can test this possibility by investigating whether Oct4λVP2 is able to repress differentiation specific gene expression in *Nanog* null cells.

5.3. "Super-pluripotency" and the Oct4-Nanog linear pathway

The Oct4λVP2 expressing cells possess characteristics of pluripotent cells. However, they are refractory to differentiation cues that I tested, and exhibit signs of LIF-independent self-renewal. This makes the Oct4 activator-only cell line a "super-pluripotent" cell line that is resistant to differentiation. This phenotype is reminiscent of *Nanog* overexpressing cells (Cavaleri and Scholer 2003; Yates and Chambers 2005). In our cell line, *Nanog* is highly upregulated and its expression correlates with Oct4λVP2 levels, strongly confirming that Oct4 directly regulates *Nanog* expression (Rodda et al., 2005). Strikingly, when these cells are grown in the presence of LIF, *Nanog* is the only key pluripotency gene that is upregulated by greater than 2-fold in comparison to wild type ES cells. Together these observations suggest that the key focal point in regulating the balance between self-renewal and differentiation is Nanog acting downstream of Oct4. To test this, careful kinetic studies on the

response of Nanog to Oct4 are necessary.

The notion that Nanog is the main effector of Oct4 and other pluripotency factors is supported by recent findings. Nanog was shown to be essential for the maintenance of pluripotency in the ICM downstream of Oct4 and other core factors and and that it is required downstream of the key factors responsible for induced pluripotency (Silva et al., 2009). Figure 5.1 illustrates the effect of Oct4λVP2 on the core pluripotency network in mouse embryonic stem cells.

5.4. C-TD of PouV proteins and importance of ERK inhibition

We found that the C-terminus of Xlpou91 is essential for self-renewal. Our bioinformatics analysis has revealed that within both the C-terminus of Oct4 and Xlpou91 lies a potential DEF site and ERK phosphorylation sites. Future work will concentrate on analysing whether Oct4 is phosphorylated by ERK and whether the DEF site, as well as the adjacent potential phosphorylation sites is involved in such regulation. A role for ERK in the regulation of Oct4 activity would be exciting, as it has been shown to be essential to maintain ES cells in the absence of serum or cytokines (Ying, Wray et al. 2008; Nichols, Silva et al. 2009). ERK inhibition also promotes the derivation of ES cells from recalcitrant mouse strains (Batlle-Morera et al., 2008) as well as rats (Buehr et al., 2008; Li et al., 2008) and increases the efficiency of reprogramming (Silva et al., 2008).

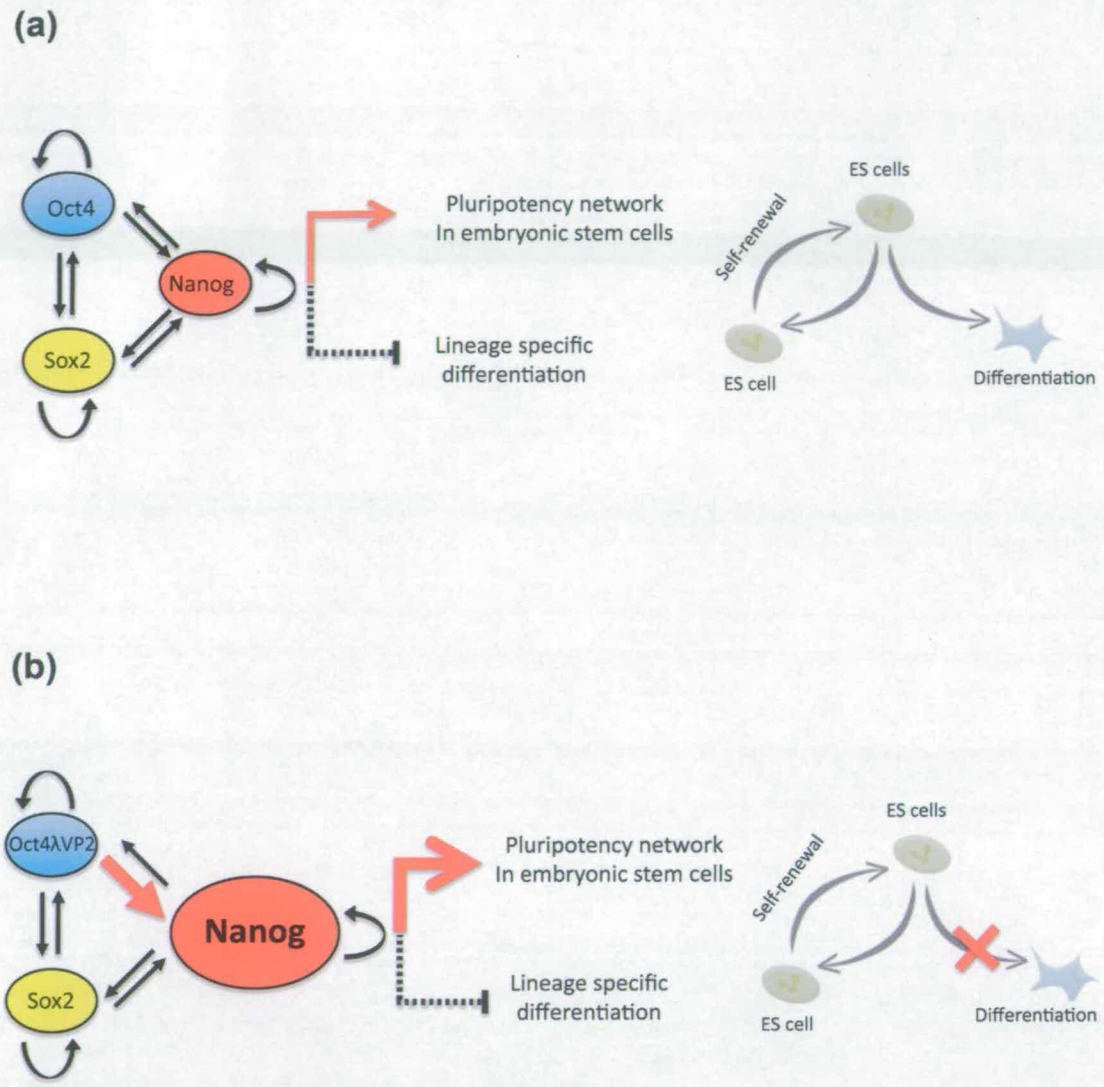


Figure 5.1-Effect of Oct4λVP2 on the core pluripotency network in mouse embryonic stem cells

(a) In mouse ES cells, the three core pluripotency factors (Oct4, Nanog and Sox2) interact with each other creating a feed forward regulatory network as well as a feedback autoregulation loop. These core factors maintain ES cells in an undifferentiated state mainly by activating regulators involved in self-renewal. This allows the cells to self-renew while retaining their ability to respond to differentiation cues.
(b) Replacing Oct4 with its activator form (Oct4λVP2) in ES cells, leads to Nanog upregulation. These high levels of Nanog seem to lock the cells in a pluripotency cycle, rendering them refractory to differentiation signals.

APPENDIX

Table 3.2- Gene ontology of upregulated genes in Oct4λVP2 cell line.

Gene ontology (GO annotations of upregulated genes in Oct4λVP2 cell line with more than 2 fold change performed by A.Sharov (NIH).

Annotation	Enrichment ratio	p	FDR	Gene symbols
extracellular space	1.654	0	0	Vnn1,Otor,Hhip,Timp4,Prlpc2,Ambp,Grin1,Cdh17,Cst9,Ly6c,Cbln1,Sez6,Tfpi,Fmo1,Emp1,Serpine1,Adipoq,Wif1,Pzp,Fxyd4,Ffar2,Wnt3a,Ptges,Crispld2,Rspo2,Gjb4,Tie1,Cd2,Klk11,Cspg5,Sparcl1,Gprc5b,Loxl4,Efnb1,Spp1,Ptgs1,Mep1b,Abp1,Pcolce2,Lgals3bp,Kdr,Apom,Frzb,Ces2,Crtac1,Prss35,Oc90,Cyp2c29,C1qtnf1,Cyp2b19,Ceacam10,Fcgrt,Nt5e,Klk1b26,Ildr1,Edn2,Csf3r,Il18r1,Mesdc,Cck,Thbd,Dct,Ang2,Angpt2,Gpr37l1,Ifna2,Ang1,B3gnt3,Htra1,Plxnc1,Lynx1,Lrig1,Nid1,Il17b,Jam4,Tex264,Pdzk1ip1,Gip,Tlr3,Tpsg1,Ddr2,Cd63,Cfc1,Wnt8b,Zp3,Agrp,Proz,Unc5cl,Tgfb1,Il6ra,Ncam1,Folr1,Rhbg,Msln,Vit,B3galnt1,Gabrr1,Tnfrsf1a,Mfge8,Lrp10,Mmp19,Lipc,Gjb5,Klk1b1,Cdh2,Tnfrsf8,Ror1,Igfbp2,Chga,Zbtb7b,Acp6,Omg,Igfbp1,Sparc,Chrnb2,Mst1,Mfrp,Lefty2,Nmu,Scg5,Frk,Casr,Gdf3,Sfrp5,Ces5,Itlna,Adam19,Dpep1,Smoc2,Aoc3,Cpvl,Mcam,Sema3b,Kitl,Pla2g10,Galns,Clgn,Klk10,Vwf,Sst,Ephb4,Tspan1,Col5a3,Gfra1,Adora2b,Gpnmb,Kcnd1,Fn1,Crhr2,Ptpre,Sqle,Tmem8,Defcrrs1,Fbln1,Cst3,Smpdl3b,Eng,Igfbp5,Sfrp1,Entpd2,Col5a1,Tff2,Ang4,C1r,Mr1,Mpo,Wnt6,Stc2,Vpreb1,Pcsk1n,Spon2,Trh,Ptprj,Vegfc,Ephx1,Pga5,
extracellular region	1.564	0	0	Vnn1,Otor,Hhip,Timp4,Sgcz,Prlpc2,Ambp,Grin1,Cdh17,Cst9,Ly6c,Cbln1,Sez6,Tfpi,Fmo1,Emp1,Serpine1,Adipoq,Wif1,Pzp,Fxyd4,Ffar2,Wnt3a,Ptges,Crispld2,Rspo2,Gjb4,Tie1,Cd2,Klk11,Cspg5,Sparcl1,Gprc5b,Il7,Loxl4,Efnb1,Spp1,Ptgs1,Mep1b,Abp1,Pcolce2,Lgals3bp,Kdr,Apom,Frzb,Ces2,Crtac1,Prss35,Oc90,Cyp2c29,C1qtnf1,Cyp2b19,Ceacam10,Fcgrt,Nt5e,Klk1b26,Saa2,Ildr1,Edn2,Csf3r,Il18r1,Mesdc2,Cck,Thbd,Dct,Ang2,Angpt2,Gpr37l1,Ifna2,Nov,Ang1,B3gnt3,Htra1,Plxnc1,Lynx1,Lrig1,Nid1,Il17b,Jam4,Tex264,Pdzk1ip1,Gip,Tlr3,Tpsg1,Ddr2,Cd63,Cfc1,Wnt8b,Zp3,Adamts4,Agrp,Proz,Unc5cl,Tgfb1,Il6ra,Ncam1,Folr1,Rhbg,Msln,Vit,B3galnt1,Gabrr1,Tnfrsf1a,Mfge8,Lrp10,Mmp19,Lipc,Gjb5,Klk1b1,Cdh2,Tnfrsf8,Ror1,Igfbp2,Chga,Zbtb7b,Acp6,Omg,Igfbp1,Sparc,Chrnb2,Fbln2,Mst1,Mfrp,Lefty2,Nmu,Scg5,Frk,Casr,Gdf3,Sfrp5,Ces5,Itlna,Chia,Adam19,Dpep1,Smoc2,Aoc3,Cpvl,Il1f8,Mcam,Sema3b,Kitl,Pla2g10,Galns,Clgn,Klk10,Vwf,Sst,Ephb4,Tspan1,Col5a3,Gfra1,Adora2b,Gpnmb,Kcnd1,Fn1,Lama1,Crhr2,Ptpre,Sqle,Tmem8,Defcrrs1,Fbln1,Cst3,Smpdl3b,Eng,Igfbp5,Sfrp1,Entpd2,Col5a1,Tff2,Ang4,C1r,Mr1,Mpo,Wnt6,Stc2,Vpreb1,Pcsk1n,Spon2,Trh,Ptprj,Vegfc,Ephx1,Pga5,
endopeptidase inhibitor activity	3.178	0	0	Serpinb1b,A2m,Timp4,Ambp,Serpinb6c,Cst9,Tfpi,Serpine1,Pzp,Pbp2,Cst13,Serpinb6b,Serpinb9b,Serpinb1a,Plxnc1,Papln,Serpinb1c,Wfikkn2,Vwf,Serpinb9,Lxn,Cst3,Pcsk1n,
protease inhibitor activity	3.151	0	0	Serpinb1b,A2m,Timp4,Ambp,Serpinb6c,Cst9,Tfpi,Serpine1,Pzp,Pbp2,Cst13,Serpinb6b,Serpinb9b,Serpinb1a,Plxnc1,Papln,Serpinb1c,Wfikkn2,Vwf,Serpinb9,Lxn,Cst3,Pcsk1n,

Annotation	Enrichment ratio	P	FDR	Gene symbols
calcium ion binding	1.909	0	0	Calm5,Pkd2l1,Myl7,S100a11,S100a14,Cdh17,Pkd1l2,Calm4,Efhb,Sparcl1,Scgn,Cacna1g,Notch4,Crtac1,Oc90,Gm467,Tesc,S100a13,Thbd,Cdh26,Nid1,Mylpf,Cabp4,Cspg2,Padi4,Proz,Pcdh8,Odz2,Megf6,Plscr2,Plek,Trpa1,Itsn1,Cdh2,Plscr1,Pclo,Chga,Sparc,Fbln2,Mst1,Pkhd1,Frem3,Smoc2,Scube3,Pla2g10,Clgn,Lcp1,Slc25a23,Fbln1,Utrn,Crb2,Tgm2,C1r,Cabp5,Anxa9,Pcdh19,Mpo,Dhx32,Heg1,Def6,
serine-type endopeptidase inhibitor activity	3.351	0	0	Serpinb1b,Ambp,Serpinb6c,Tfpi,Serpine1,Pzp,Pbp2,Serpinb6b,Serpinb9b,Serpinb1a,Plxnc1,Papln,Serpinb1c,Wfikkn2,Vwf,Serpinb9,Pcsk1n,
enzyme inhibitor activity	2.558	0	0	Serpinb1b,A2m,Timp4,Ambp,Serpinb6c,Ppp1r14d,Cst9,Tfpi,Serpine1,Pzp,Pbp2,Cst13,Serpinb6b,Serpinb9b,Serpinb1a,Plxnc1,Papln,Serpinb1c,Wfikkn2,Vwf,Serpinb9,Lxn,Cst3,Ppp1r1b,Pcsk1n,
alpha-type channel activity	2.067	0	0	Grid2,Aqp9,Pkd2l1,Grin1,Pkd1l2,Fxyd4,Gjb4,Aqp3,Gjb3,Cacna1g,Kcna5,Kcnc4,Scn3a,Clic3,Clca1,Grin2d,Scnn1b,Kcnn3,Ryr3,Itpr3,Kctd8,Aqp7,Gabrr1,Gjb5,Trpa1,Gjc1,Chrnb2,Clcnka,P2rx5,Kctd1,Gja7,Accn3,Trpm6,Kcns3,Kctd15,Kcnd1,Kcnj12,Cnga3,Kcnu1,
transporter activity	1.536	0	0	Grid2,Abcb11,Aqp9,Pkd2l1,Ambp,Grin1,Cdh17,Abca4,Slc6a7,Fmo1,Pkd1l2,Slc16a9,Fxyd4,Slc4a1,Gjb4,Aqp3,Slc16a5,Gjb3,Slco2b1,Rbp1,Abp1,Atp12a,Cacna1g,Apom,Notch4,Fabp3,Atp6v0a4,Kcna5,Saa2,Abcc3,Nrxn3,Slc15a1,Atp6v1c2,Slc5a1,Sfxn3,Slc30a2,Kcnc4,Slc4a2,Slc6a13,Scn3a,Cygb,Robo4,Lynx1,Slc28a1,Clic3,Slc5a4a,Cd63,Slc25a1,Clca1,Grin2d,Scnn1b,Kcnn3,Ryr3,Atp8a2,Slc7a9,Itpr3,Slc6a4,Nxf2,Folr1,Rhbg,Kctd8,Aqp7,Slc23a1,Slc23a2,Slc13a5,Gabrr1,Lipc,Gjb5,Trpa1,Gjc1,Crabp2,Slc25a20,Chrnb2,Ap3m2,Fabp1,Clcnka,P2rx5,Rlbp1,Kctd1,Gja7,Snx1,Slc39a4,Accn3,Abcd1,Slco2a1,Trpm6,Sv2c,Kcns3,Kctd15,Kcnd1,Ucp3,Slc12a7,Slc25a23,Slc5a2,Slc16a4,Kcnj12,Slc5a4b,Cnga3,Slc36a3,Kcnu1,Slc4a11,Pgm1,Uqcrc1,Atp1b3,
prostaglandin metabolism	9.238	0	0	Ptges,Ptgs1,Hpgd,Tnfrsf1a,
intracellular ligand-gated ion channel activity	9.238	0	0	Clca1,Ryr3,Itpr3,Cnga3,
prostanoid metabolism	9.238	0	0	Ptges,Ptgs1,Hpgd,Tnfrsf1a,
plasma membrane	1.516	0	0	Abcb11,Aqp9,Hhip,Sgcz,Drp2,Cdh17,Ly6c,Abca4,Slc6a7,Ly6a,Entpd3,Lim2,Kit,Edg2,Slc4a1,Gjb4,Cd2,Sstr2,Blnk,Aqp3,Thy1,Efnb1,Gjb3,Cd37,Cacna1g,Cnnm1,Ptprb,Ggtla1,Crtac1,Kcna5,Fcgrt,Fcgr2b,Ptch1,Abcc3,Odz1,Slc5a1,Il18r1,Kcnc4,Slc4a2,Thbd,Slc6a13,Robo4,Cd3d,Cryab,Lynx1,Cd82,Icam2,Jam4,Prkcq,Itgb3,Slc25a1,Clca1,Ap3b2,Il6ra,Slc6a4,Pcdh8,Ap1g2,Ncam1,Odz2,Rhbg,Zap70,Kctd8,Aqp7,Slc23a1,Slc13a5,Gabrr1,Tnfrsf1a,Mfge8,Lrp10,Gjb5,Gjc1,Cdh2,Ror1,Ptprcap,P2rx5,Cd300d,Kctd1,Gja7,Slc39a4,Mcam,Accn3,Kitl,Nf2,Slco2a1,Nfam1,Kcns3,Kctd15,Lcp1,Gpnmb,Kcnd1,Slc12a7,Il10rb,Ptpre,Tmem8,Rapsn,Utrn,Kcnj12,Entpd2,Mr1,Tln1,Kcnu1,Vpreb1,Ptprj,
regulation of body fluids	3.943	0	0	Nfe2,Tfpi,Thbd,Proz,Mst1,Rab27a,Clcnka,Vwf,Il10rb,Entpd2,
channel or pore class transporter activity	1.995	0	0	Grid2,Aqp9,Pkd2l1,Grin1,Pkd1l2,Fxyd4,Gjb4,Aqp3,Gjb3,Cacna1g,Kcna5,Kcnc4,Scn3a,Clic3,Clca1,Grin2d,Scnn1b,Kcnn3,Ryr3,Itpr3,Kctd8,Aqp7,Gabrr1,Gjb5,Trpa1,Gjc1,Chrnb2,Clcnka,P2rx5,Kctd1,Gja7,Accn3,Trpm6,Kcns3,Kctd15,Kcnd1,Kcnj12,Cnga3,Kcnu1,
angiogenesis	3.403	0	0.001	Serpine1,Tie1,Kdr,Angpt2,Robo4,Ang1,Mmp19,Edg1,Crhr2,Eng,Epas1,Vegfc,

Annotation	Enrichment ratio	p	FDR	Gene symbols
blood vessel morphogenesis	3.233	0	0.001	Serpine1,Tie1,Kdr,Angpt2,Robo4,Ang1,Mmp19,Edg1,Crhr2,Eng,Epas1,Ptprj,Vegfc,
protein polymerization	4.715	0	0.001	Tubb3,Tubb2a,Gas7,Tuba3,Tubal3,Ttll3,Tubb2b,
connexon complex	6.218	0	0.001	Gjb4,Gjb3,Gjb5,Gjc1,Gja7,
connexon channel activity	6.218	0	0.001	Gjb4,Gjb3,Gjb5,Gjc1,Gja7,
gap-junction forming channel activity	6.218	0	0.001	Gjb4,Gjb3,Gjb5,Gjc1,Gja7,
structural constituent of eye lens	6.218	0	0.001	Lim2,Cryge,Crygf,Cryab,Crygb,
hemostasis	3.829	0	0.002	Nfe2,Tfpi,Thbd,Proz,Mst1,Rab27a,Vwf,Il10rb,Entpd2,
cell growth	3.355	0.0001	0.002	Emp1,Fhl1,Nov,Htra1,Tgfb1,Igfbp2,Igfbp1,Lefty2,Socs7,Creg1,Igfbp5,
organ development	1.627	0.0001	0.002	Otor,Hhip,Myl7,Tnnt2,Serpine1,Hesx1,Lim2,Kit,En1,Phex,Wnt3a,Efhb,Tie1,Myom2,Il7,Spp1,Pax7,Kdr,Ptch1,Mybpc3,Fhl1,Atoh1,Slc5a1,Angpt2,Hoxb6,Robo4,Ang1,Cryab,Crygb,Mylpf,Id3,Sfpi1,Wnt8b,Adamts4,Tgfb1,Hoxc12,Zap70,Mmp19,Hspb2,Chat,Mitf,Edg1,Rpgrip1,Pkhd1,Cxcr4,Adam19,Gadd45g,Hmgb3,Nfam1,Pax6,Sox9,Myod1,Crhr2,Utrn,Eng,Wnt6,Vpreb1,Vdr,Csrp3,Epas1,Smyd1,Ptprj,Vegfc,
icosanoid metabolism	4.85	0.0001	0.003	Alox12e,Ptges,Ptgs1,Ggtla1,Hpgd,Tnfrsf1a,
regulation of cell size	3.175	0.0002	0.004	Emp1,Fhl1,Nov,Htra1,Tgfb1,Igfbp2,Igfbp1,Lefty2,Socs7,Creg1,Igfbp5,
cell fraction	1.94	0.0002	0.004	Grid2,Grin1,Cbln1,Fmo1,Calm4,Ptges,Sstr2,Ptgs1,Cacna1g,Cyp2c29,Cyp2b19,Fads2,Cd38,Cryab,Lnpep,Lynx1,Clic3,Dad1,A4galt,Folr1,Dgat2,Ulk1,Hspb2,Chrnb2,Drd1a,Rac2,Dpep1,Cyp2c37,Cpne6,Utrn,H6pd,Ephx1,Pga5,
regulation of lymphocyte activation	3.804	0.0002	0.004	Il7,Efnb1,Fcgr2b,Prkcq,Tgfb1,Zap70,Hmgb3,Nfam1,
regulation of cell activation	3.804	0.0002	0.004	Il7,Efnb1,Fcgr2b,Prkcq,Tgfb1,Zap70,Hmgb3,Nfam1,
response to wounding	2.209	0.0003	0.005	Tfpi,Ggtla1,Fcgr2b,Csf3r,Thbd,Rel,Il17b,Prkcq,Tlr3,Proz,Tgfb1,Tnfrsf1a,Mst1,Rab27a,Ccr4,Il1f8,Gadd45g,Vwf,Lcp1,Fn1,Il10rb,Entpd2,
regulation of cell migration	4.042	0.0003	0.005	Tie1,Robo4,Itgb3,Cxcr4,Sst,Pax6,Lama1,
blood vessel development	2.765	0.0003	0.006	Serpine1,Tie1,Kdr,Angpt2,Robo4,Ang1,Mmp19,Edg1,Crhr2,Eng,Epas1,Ptprj,Vegfc,
striated muscle thick filament	8.084	0.0004	0.006	Myom2,Mybpc3,Obscn,
A band	8.084	0.0004	0.006	Myom2,Mybpc3,Obscn,
photoreceptor cell development	8.084	0.0004	0.006	Nr2e3,Rpgrip1,Pax6,

Annotation	Enrichment ratio	P	FDR	Gene symbols
response to external stimulus	1.961	0.0004	0.006	Pdcl2,Tfpi,Kit,Ggtla1,Fcgr2b,Csf3r,Thbd,Rel,Il17b,Prkcq,Tlr3,Proz,Tgfb1,Tnfrsf1a,Mst1,Rab27a,Ccr4,Rac2,Cxcr4,Cxcr3,Il1f8,Accn3,Gadd45g,Vwf,Lcp1,Fn1,Il10rb,Utrn,Entpd2,
inorganic anion exchanger activity	8.084	0.0004	0.006	Slc4a1,Slc4a2,Slc4a11,
regulation of B cell activation	5.052	0.0004	0.006	Il7,Fcgr2b,Tgfb1,Hmgb3,Nfam1,
insulin-like growth factor binding	5.052	0.0004	0.005	Nov,Htra1,Igfbp2,Igfbp1,Igfbp5,
gap junction	5.052	0.0004	0.005	Gjb4,Gjb3,Gjb5,Gjc1,Gja7,
lymphocyte proliferation	3.902	0.0004	0.006	Il7,Efnb1,Fcgr2b,Prkcq,Zap70,Cxcr4,Ccnd3,
regulation of cell motility	3.902	0.0004	0.006	Tie1,Robo4,Itgb3,Cxcr4,Sst,Pax6,Lama1,
vasculature development	2.694	0.0005	0.007	Serpine1,Tie1,Kdr,Angpt2,Robo4,Ang1,Mmp19,Edg1,Crhr2,Eng,Epas1,Ptprj,Vegfc,
UDP-galactosyltransferase activity	5.879	0.0005	0.007	Abo,A4galt,B3galnt1,Wdfy3,
regulation of lymphocyte differentiation	5.879	0.0005	0.006	Il7,Zap70,Hmgb3,Nfam1,
membrane fraction	1.929	0.0005	0.007	Grid2,Grin1,Cbln1,Fmo1,Ptges,Sstr2,Ptgs1,Cacna1g,Cyp2c29,Cyp2b19,Fads2,Cd38,Lnpep,Lynx1,Clic3,Dad1,A4galt,Folr1,Dgat2,Ulk1,Chrnb2,Drd1a,Rac2,Dpep1,Cyp2c37,Cpne6,Utrn,H6pd,Ephx1,
blood coagulation	3.495	0.0006	0.007	Tfpi,Thbd,Proz,Mst1,Rab27a,Vwf,Il10rb,Entpd2,
morphogenesis	1.599	0.0006	0.007	Hhip,Emp1,Serpine1,En1,Wnt3a,Tie1,Nr2e3,Efnb1,Pax7,Kdr,Ptch1,Fhl1,Atoh1,Cck,Angpt2,Hoxb6,Nov,Robo4,Ang1,Htra1,Chl1,Id3,Wnt8b,Tgfb1,Hoxc12,Gas7,Pcdh8,Ulk1,Mmp19,Igfbp2,Igfbp1,Edg1,Lefty2,Rpgrip1,Rac2,Adam19,Socs7,Plekhc1,Lhx1,Pax6,Wasl,Gfra1,Sox9,Lama1,Crhr2,Eng,Creg1,Igfbp5,Snai1,Wnt6,Vdr,Epas1,Smyd1,Ptprj,Vegfc,
response to biotic stimulus	1.647	0.0006	0.007	Oas1d,Ly6c,Oas1b,Cd2,Blnk,Il7,Gbp4,Efnb1,Oas2,Ggtla1,Fcgrt,Fcgr2b,Saa2,Csf3r,Oas1c,Ifna2,Oas1e,Rel,Il17b,Prkcq,Tlr3,Sfpi1,Tnfrsf14,Tgfb1,Zap70,Tnfrsf1a,Oas1h,Tnfsf18,Ccr4,Tcea3,Cxcr4,Itlna,Cxcr3,Il1f8,Cfp,Gadd45g,Hmgb3,Nfam1,Wasl,Fn1,Ccnd3,Defcr-rs1,Ang4,Oasl2,C1r,Mr1,Vpreb1,Spon2,
wound healing	3.163	0.0008	0.009	Tfpi,Thbd,Proz,Mst1,Rab27a,Vwf,Fn1,Il10rb,Entpd2,
regulation of locomotion	3.65	0.0009	0.01	Tie1,Robo4,Itgb3,Cxcr4,Sst,Pax6,Lama1,
coagulation	3.316	0.001	0.011	Tfpi,Thbd,Proz,Mst1,Rab27a,Vwf,Il10rb,Entpd2,
positive regulation of lymphocyte proliferation	5.389	0.001	0.011	Il7,Efnb1,Prkcq,Zap70,
response to drug	5.389	0.001	0.011	Phf11,D14Ertd668e,Abp1,Drd1a,

Annotation	Enrichment ratio	p	FDR	Gene symbols
defense response	1.634	0.001	0.011	Oas1d,Ly6c,Oas1b,Cd2,Blnk,Il7,Gbp4,Efnb1,Oas2,Ggtla1,Fcgrt,Fcgr2b,Saa2,Csf3r,Oas1c,Ifna2,Oas1e,Rel,Il17b,Prkcq,Tlr3,Sfpi1,Tnfrsf14,Tgfb1,Zap70,Tnfrsf1a,Oas1h,Tnfsf18,Ccr4,Tcea3,Cxcr4,Cxcr3,Il1f8,Cfp,Gadd45g,Hmgb3,Nfam1,Fn1,Ccnd3,Defcr-rs1,Ang4,Oasl2,C1r,Mr1,Vpreb1,Spon2,
tubulin-tyrosine ligase activity	6.929	0.001	0.011	Ttll6,Ttll3,Ttl,
regulation of angiogenesis	6.929	0.001	0.011	Serpine1,Tie1,Crhr2,
water channel activity	6.929	0.001	0.01	Aqp9,Aqp3,Aqp7,
immune response	1.699	0.0012	0.011	Oas1d,Oas1b,Il7,Gbp4,Efnb1,Oas2,Ggtla1,Fcgrt,Fcgr2b,Saa2,Csf3r,Oas1c,Oas1e,Rel,Il17b,Prkcq,Tlr3,Sfpi1,Tnfrsf14,Tgfb1,Zap70,Tnfrsf1a,Oas1h,Tnfsf18,Ccr4,Cxcr4,Il1f8,Cfp,Gadd45g,Hmgb3,Nfam1,Fn1,Ccnd3,Ang4,Oasl2,C1r,Mr1,Vpreb1,Spon2,
ion transporter activity	1.566	0.0013	0.012	Grid2,Pkd2l1,Grin1,Abca4,Slc6a7,Pkd1l2,Fxyd4,Slc4a1,Atp12a,Cacna1g,Notch4,Atp6v0a4,Kcna5,Atp6v1c2,Slc5a1,Sfxn3,Slc30a2,Kcnc4,Slc4a2,Slc6a13,Scn3a,Slc28a1,Clic3,Clca1,Grin2d,Scnn1b,Kcnn3,Ryr3,Atp8a2,Itpr3,Slc6a4,Rhbg,Kctd8,Slc23a2,Gabrr1,Trpa1,Chrnb2,Clcnka,P2rx5,Kctd1,Slc39a4,Accn3,Trpm6,Kcns3,Kctd15,Kcnd1,Slc12a7,Kcnj12,Cnga3,Kcnu1,Slc4a11,Uqcrc1,Atp1b3,
oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	2.502	0.0013	0.012	Akr1c18,Hsd17b2,Hpgd,Ldhal6b,Idh2,Rnasel,Akr1c13,Akr1b7,Me3,Tdh,Gpd1,H6pd,Adh1,
extracellular matrix	1.959	0.0013	0.012	Timp4,Sgcz,Wnt3a,Sparcl1,Oc90,Adamts6,Nid1,Papln,Zp3,Adamts4,Tgfb1,Mmp19,Sparc,Fbln2,Adam19,Smoc2,Vwf,Col5a3,Fn1,Lama1,Fbln1,Entpd2,Col5a1,Spon2,
growth factor binding	3.429	0.0015	0.014	Il1rl2,Nov,Htra1,Igfbp2,Igfbp1,Il10rb,Igfbp5,
growth	2.382	0.0016	0.015	Emp1,Ptch1,Fhl1,Nov,Htra1,Tgfb1,Igfbp2,Igfbp1,Lefty2,Gdf3,Socs7,Lgmn,Creg1,Igfbp5,
enzyme linked receptor protein signaling pathway	2.104	0.0017	0.015	Hhip,Kit,Rspo2,Clnk,Kdr,Ptprb,Smad6,Angpt2,Htra1,Ddr2,Erbb3,Tgfb1,Ptprcap,Wfikkn2,Ephb4,Gfra1,Ptpre,Eng,Ptprj,
ligand-gated ion channel activity	2.552	0.0017	0.014	Grid2,Grin1,Clca1,Grin2d,Scnn1b,Ryr3,Itpr3,Gabrr1,Chrnb2,P2rx5,Accn3,Cnga3,
lymphocyte activation	2.552	0.0017	0.014	Il7,Efnb1,Fcgr2b,Prkcq,Sfpi1,Tgfb1,Zap70,Cxcr4,Gadd45g,Hmgb3,Nfam1,Ccnd3,
ion transport	1.523	0.0017	0.014	Grid2,Pkd2l1,Grin1,C1ql3,Adipoq,Pkd1l2,Lim2,Fxyd4,Slc4a1,Slco2b1,Atp12a,Cacna1g,Atp6v0a4,C1qtnf1,Kcna5,Atp6v1c2,Slc5a1,Sfxn3,Slc30a2,Kcnc4,Slc4a2,Scn3a,Clic3,Clca1,Grin2d,Scnn1b,Kcnn3,Ryr3,Atp8a2,Itpr3,Kctd8,Slc23a1,Slc23a2,Slc13a5,Gabrr1,Trpa1,Tst,Chrnb2,Clcnka,P2rx5,Kctd1,Slc39a4,Accn3,Trpm6,Col5a3,Kcns3,Kctd15,Kcnd1,Slc12a7,Slc5a2,Kcnj12,Col5a1,Cnga3,Kcnu1,Vdr,Slc4a11,Atp1b3,
transmembrane receptor protein tyrosine phosphatase signaling pathway	4.974	0.0018	0.014	Ptprb,Ptprcap,Ptpre,Ptprj,

Annotation	Enrichment ratio	p	FDR	Gene symbols
chymotrypsin activity	2.287	0.0019	0.015	Tmprss7,Klk11,Tmprss11d,Prss35,Klk1b26,Htra1,Tpsg1,Proz,Klk1b1,Tmprss5,Mst1,Klk1b24,Klk10,Klk13,C1r,
ion channel activity	1.764	0.002	0.016	Grid2,Pkd2l1,Grin1,Pkd1l2,Fxyd4,Cacna1g,Kcna5,Kcnc4,Scn3a,Clic3,Clca1,Grin2d,Scnn1b,Kcnn3,Ryr3,Itpr3,Kctd8,Gabrr1,Trpa1,Chrnb2,Clcnka,P2rx5,Kctd1,Accn3,Trpm6,Kcns3,Kctd15,Kcnd1,Kcnj12,Cnga3,Kcnu1,
muscle development	2.333	0.0021	0.017	Myl7,Tnnt2,Efhb,Myom2,Mybpc3,Fhl1,Cryab,Mylpf,Tgfb1,Hspb2,Chat,Myod1,Utrn,Csrp3,
fatty acid metabolism	2.266	0.0021	0.016	Alox12e,Adipoq,Ptges,Scd1,Ptgs1,Acoxl,Ggtla1,Fads2,Hpgd,Tnfrsf1a,Slc27a3,Fads1,Ucp3,Prkag1,Scap,
water transporter activity	6.063	0.0023	0.018	Aqp9,Aqp3,Aqp7,
carbon-nitrogen ligase activity, with glutamine as amido-N-donor	6.063	0.0023	0.017	Cps1,Gdpd2,Gpnmb,
anion exchanger activity	6.063	0.0023	0.017	Slc4a1,Slc4a2,Slc4a11,
water transport	6.063	0.0023	0.017	Aqp9,Aqp3,Aqp7,
anion:anion antiporter activity	6.063	0.0023	0.017	Slc4a1,Slc4a2,Slc4a11,
pregnancy	6.063	0.0023	0.017	Thbd,Mst1,Sult1e1,
bicarbonate transporter activity	6.063	0.0023	0.017	Slc4a1,Slc4a2,Slc4a11,
fluid transport	6.063	0.0023	0.017	Aqp9,Aqp3,Aqp7,
eye development (sensu Vertebrata)	4.042	0.0024	0.017	Lim2,Cryab,Crygb,Mitf,Pax6,
regulation of lymphocyte proliferation	4.042	0.0024	0.017	Il7,Efnb1,Fcgr2b,Prkcq,Zap70,
galactosyltransferase activity	4.042	0.0024	0.017	Abo,B3gnt3,A4galt,B3galnt1,Wdfy3,
cation transporter activity	1.601	0.0024	0.017	Pkd2l1,Grin1,Abca4,Slc6a7,Pkd1l2,Atp12a,Cacna1g,Notch4,Atp6v0a4,Kcna5,Atp6v1c2,Slc5a1,Sfxn3,Slc30a2,Kcnc4,Slc6a13,Scn3a,Slc28a1,Scnn1b,Kcnn3,Ryr3,Atp8a2,Itpr3,Slc6a4,Rhbg,Kctd8,Slc23a2,Trpa1,Chrnb2,P2rx5,Kctd1,Slc39a4,Accn3,Trpm6,Kcns3,Kctd15,Kcnd1,Slc12a7,Kcnj12,Cnga3,Kcnu1,Uqcrc1,Atp1b3,
nucleotide catabolism	4.619	0.0029	0.02	Entpd3,Nt5e,Upp1,Entpd2,
transmembrane receptor protein tyrosine kinase activity	2.939	0.0031	0.022	Kit,Rspo2,Kdr,Ddr2,Erbb3,Ncam1,Ror1,Ephb4,

Annotation	Enrichment ratio	P	FDR	Gene symbols
cell adhesion	1.62	0.0032	0.021	Cdh17,Cd2,Siglecf,Spp1,Crtac1,Mybpc3,Csf3r,Ctnnal1,Cdh26,Nid1,Chl1,Icam2,Jam4,Itgb3,Ddr2,Cspg2,Pcdh8,Ncam1,Mfge8,Cdh2,Omg,Edg1,Frem3,Aoc3,Prtg,Mcam,Kitl,Crim2,Plekhc1,Vwf,Col5a3,Gpnmb,Fn1,Lama1,Tmem8,Eng,Col5a1,Pcdh19,Spon2,
cytokine binding	2.745	0.0034	0.022	Il1rl2,Il18r1,Il6ra,Tnfrsf1a,Tnfrsf8,Ccr4,Cxcr4,Cxcr3,Il10rb,
T cell proliferation	3.849	0.0034	0.023	Efnb1,Prkcq,Zap70,Cxcr4,Ccnd3,
interaction between organisms	3.849	0.0034	0.022	Edn2,Thbd,Mst1,Sult1e1,Ppp1r1b,
cation channel activity	1.839	0.0037	0.024	Pkd2l1,Grin1,Pkd1l2,Cacna1g,Kcna5,Kcnc4,Scn3a,Scnn1b,Kcnn3,Ryr3,Itpr3,Kctd8,Trpa1,Chrnb2,P2rx5,Kctd1,Accn3,Trpm6,Kcns3,Kctd15,Kcnd1,Kcnj12,Cnga3,Kcnu1,
oxidoreductase activity, acting on CH-OH group of donors	2.284	0.004	0.025	Akr1c18,Hsd17b2,Hpgd,Ldhal6b,Idh2,Rnasel,Akr1c13,Akr1b7,Me3,Tdh,Gpd1,H6pd,Adh1,
organic acid metabolism	1.645	0.004	0.025	Alox12e,Adipoq,Ptges,Scd1,Gjb3,Ptgs1,Creg2,Slc45a2,Acoxl,Ggtla1,Fads2,Aspa,Cps1,Dct,Nr5a2,Uroc1,Baat,Hpgd,Folr1,Ktn1,Slc23a2,Tnfrsf1a,Idh2,Slc27a3,Fads1,Hal,Kitl,Tdo2,Ucp3,Glul,Me3,Gpd1,Prkag1,Scap,Cdo1,
carboxylic acid metabolism	1.645	0.004	0.025	Alox12e,Adipoq,Ptges,Scd1,Gjb3,Ptgs1,Creg2,Slc45a2,Acoxl,Ggtla1,Fads2,Aspa,Cps1,Dct,Nr5a2,Uroc1,Baat,Hpgd,Folr1,Ktn1,Slc23a2,Tnfrsf1a,Idh2,Slc27a3,Fads1,Hal,Kitl,Tdo2,Ucp3,Glul,Me3,Gpd1,Prkag1,Scap,Cdo1,
tissue kallikrein activity	5.389	0.0045	0.027	Klk1b26,Klk1b1,Klk1b24,
photoreceptor cell differentiation	5.389	0.0045	0.027	Nr2e3,Rpgrip1,Pax6,
physiological interaction between organisms	5.389	0.0045	0.027	Thbd,Mst1,Sult1e1,
extracellular matrix (sensu Metazoa)	1.862	0.0046	0.028	Timp4,Sgcz,Wnt3a,Sparcl1,Oc90,Nid1,Zp3,Adamts4,Tgfb1,Mmp19,Sparc,Fbln2,Adam19,Smoc2,Vwf,Col5a3,Fn1,Lama1,Fbln1,Entpd2,Col5a1,Spon2,
serine-type peptidase activity	1.886	0.0048	0.029	Tmprss7,Serpine1,Klk11,Tmprss11d,Prss35,Klk1b26,Odz1,Htra1,Tpsg1,Rimbp2,Proz,Klk1b1,Tmprss5,Mst1,Cpvl,Klk1b24,Clgn,Klk10,Klk13,C1r,Prep,
serine-type endopeptidase activity	1.944	0.005	0.03	Tmprss7,Serpine1,Klk11,Tmprss11d,Prss35,Klk1b26,Odz1,Htra1,Tpsg1,Proz,Klk1b1,Tmprss5,Mst1,Klk1b24,Clgn,Klk10,Klk13,C1r,Prep,
phospholipid binding	2.978	0.0052	0.031	Hip1,Plekha2,Sytl4,Mfge8,Pclo,Cpne6,Anxa9,
heparin binding	2.978	0.0052	0.031	Abp1,Pcolce2,Ncam1,Lipc,Gpnmb,Fn1,Col5a1,
transmembrane receptor protein kinase activity	2.598	0.0056	0.033	Kit,Rspo2,Kdr,Ly6g6e,Ddr2,Erbb3,Ncam1,Ror1,Ephb4,

Annotation	Enrichment ratio	p	FDR	Gene symbols
cellular morphogenesis	1.788	0.0057	0.033	Emp1,Wnt3a,Nr2e3,Efnb1,Fhl1,Cck,Nov,Htra1,Chl1,Tgfb1,Gas7,Ulk1,Igfbp2,Igfbp1,Lefty2,Rpgrip1,Rac2,Socs7,Plekhc1,Pax6,Wasl,Lama1,Creg1,Igfbp5,
immune cell activation	2.282	0.0057	0.033	Il7,Efnb1,Fcgr2b,Prkcq,Sfpi1,Tgfb1,Zap70,Cxcr4,Gadd45g,Hmgb3,Nfam1,Ccnd3,
cell activation	2.256	0.0064	0.037	Il7,Efnb1,Fcgr2b,Prkcq,Sfpi1,Tgfb1,Zap70,Cxcr4,Gadd45g,Hmgb3,Nfam1,Ccnd3,
glutathione transferase activity	4.042	0.0067	0.038	Gsta3,Gstp1,Gsta2,Gsto1,
cellular lipid metabolism	1.632	0.0069	0.039	Alox12e,Abca4,St8sia6,Pla2g4e,Adipoq,Ptges,Scd1,Akr1c18,Hsd17b2,Ptgs1,Acoxl,Ggtla1,Fabp3,Fads2,Nr5a2,Baat,Hpgd,A4galt,Dgat2,Tnfrsf1a,Slc27a3,Fads1,Sult1e1,Osbpl10,Akr1b7,Ucp3,Pla2g4d,Pmvk,Pip5k2b,Prkag1,Scap,Adh1,
trypsin activity	2.055	0.0069	0.039	Tmprss7,Klk11,Tmprss11d,Prss35,Klk1b26,Htra1,Tpsg1,Proz,Klk1b1,Tmprss5,Mst1,Klk1b24,Klk10,Klk13,C1r,
cell-cell signaling	1.835	0.007	0.038	Grid2,Grin1,Wnt3a,Gjb4,Gjb3,Grap,Naalad2,Nrxn3,Lynx1,Wnt8b,Gabrr1,Chat,Gjb5,Pclo,Chrnb2,Drd1a,Gja7,Sv2c,Dlg7,Rapsn,Wnt6,
chloride transporter activity	4.85	0.0079	0.043	Slc4a1,Slc4a2,Slc4a11,
symporter activity	2.377	0.0081	0.044	Slc6a7,Slc15a1,Slc5a1,Slc6a13,Slc28a1,Slc6a4,Slc23a1,Slc23a2,Slc12a7,Slc5a2,
cell surface	2.28	0.0082	0.044	Hhip,Ly6c,Ly6a,Kit,Thy1,Fcgr2b,Robo4,Jam4,Itgb3,Mfge8,Nfam1,
T cell activation	2.76	0.0094	0.05	Il7,Efnb1,Prkcq,Zap70,Cxcr4,Gadd45g,Ccnd3,
lipid metabolism	1.532	0.0101	0.053	Alox12e,Abca4,St8sia6,Pla2g4e,Adipoq,Ptges,Scd1,Akr1c18,Hsd17b2,Ptgs1,Acoxl,Ggtla1,Fabp3,Oc90,Fads2,Aldh1a3,Cidea,Nr5a2,Baat,Hpgd,A4galt,Dgat2,Tnfrsf1a,Lrp10,Lipc,Slc27a3,Fads1,Sult1e1,Osbpl10,Pla2g10,Akr1b7,Ucp3,Pla2g4d,Pmvk,Pip5k2b,Prkag1,Scap,Adh1,
regulation of immune response	2.425	0.0102	0.054	Il7,Efnb1,Fcgr2b,Rel,Prkcq,Tgfb1,Zap70,Hmgb3,Nfam1,
metal ion transport	1.611	0.0124	0.064	Grin1,Lim2,Atp12a,Cacna1g,Kcna5,Slc5a1,Sfxn3,Kcnc4,Scnn1b,Kcnn3,Itpr3,Kctd8,Slc23a1,Slc23a2,Slc13a5,Kctd1,Slc39a4,Accn3,Trpm6,Kcns3,Kctd15,Kcnd1,Slc12a7,Slc5a2,Kcnj12,Cnga3,Kcnu1,Vdr,Atp1b3,
positive regulation of T cell proliferation	4.409	0.0126	0.066	Efnb1,Prkcq,Zap70,
pigmentation during development	4.409	0.0126	0.065	Kit,Slc45a2,Dct,
pigmentation	4.409	0.0126	0.065	Kit,Slc45a2,Dct,
cation transport	1.509	0.013	0.066	Pkd2l1,Grin1,Pkd1l2,Lim2,Atp12a,Cacna1g,Atp6v0a4,Kcna5,Atp6v1c2,Slc5a1,Sfxn3,Slc30a2,Kcnc4,Scn3a,Scnn1b,Kcnn3,Ryr3,Atp8a2,Itpr3,Kctd8,Slc23a1,Slc23a2,Slc13a5,Trpa1,Kctd1,Slc39a4,Accn3,Trpm6,Kcns3,Kctd15,Kcnd1,Slc12a7,Slc5a2,Kcnj12,Cnga3,Kcnu1,Vdr,Atp1b3,
positive regulation of transferase activity	3.592	0.0132	0.067	Map3k6,Gadd45g,Nrk,Irak2,

Annotation	Enrichment ratio	p	FDR	Gene symbols
secretory granule	3.592	0.0132	0.066	Sytl4,Scg5,Ang4,Pcsk1n,
positive regulation of protein kinase activity	3.592	0.0132	0.066	Map3k6,Gadd45g,Nrk,Irak2,
fatty acid biosynthesis	2.632	0.0134	0.066	Alox12e,Scd1,Ptgs1,Ggtla1,Fads2,Fads1,Prkag1,
positive regulation of cell activation	3.109	0.0142	0.069	Il7,Efnb1,Prkcq,Tgfb1,Zap70,
positive regulation of lymphocyte activation	3.109	0.0142	0.069	Il7,Efnb1,Prkcq,Tgfb1,Zap70,
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	2.064	0.0153	0.074	Fmo1,Scd1,Cyp2a4,Ptgs1,Cyp2c29,Cyp2b19,Fads2,Nos1,Fads1,Cyp24a1,Cyp2c37,Sqle,
polysaccharide binding	2.309	0.0153	0.074	Abp1,Pcolce2,Cspg2,Ncam1,Lipc,Chia,Gpnmb,Fn1,Col5a1,
solute:sodium symporter activity	2.771	0.0159	0.077	Slc6a7,Slc5a1,Slc6a13,Slc28a1,Slc6a4,Slc23a2,
glycosaminoglycan binding	2.395	0.0171	0.081	Abp1,Pcolce2,Cspg2,Ncam1,Lipc,Gpnmb,Fn1,Col5a1,
neuropeptide signaling pathway	2.273	0.0174	0.082	Pamci,Pkd1l2,Gpr116,Sstr2,Agrp,Mchr1,Nmu,Scg5,Rassf2,
carbohydrate binding	1.685	0.0175	0.081	Pkd1l2,Siglecf,Clec4a3,Abp1,Pcolce2,Ptprb,Clec4a1,Thbd,Cd209d,Lman1l,Cspg2,Asb2,Asgr2,Ncam1,Lipc,Itlna,Chia,Fbxo2,Gpnmb,Fn1,Ccnd3,Col5a1,
aromatic amino acid family metabolism	3.403	0.0177	0.081	Slc45a2,Dct,Ktn1,Tdo2,
organ morphogenesis	1.549	0.0193	0.087	Hhip,Serpine1,Wnt3a,Tie1,Pax7,Kdr,Ptch1,Atoh1,Angpt2,Hoxb6,Robo4,Ang1,Id3,Wnt8b,Tgfb1,Hoxc12,Mmp19,Edg1,Rpgrip1,Adam19,Pax6,Sox9,Crhr2,Eng,Wnt6,Vdr,Epas1,Smyd1,Ptprj,Vegfc,
external side of plasma membrane	2.351	0.0195	0.088	Ly6c,Ly6a,Kit,Thy1,Fcgr2b,Robo4,Itgb3,Mfge8,
pattern binding	2.238	0.0196	0.088	Abp1,Pcolce2,Cspg2,Ncam1,Lipc,Chia,Gpnmb,Fn1,Col5a1,
regulation of growth	2.238	0.0196	0.088	Ptch1,Nov,Htra1,Igfbp2,Igfbp1,Socs7,Lgmn,Creg1,Igfbp5,
carboxylic acid biosynthesis	2.46	0.0215	0.095	Alox12e,Scd1,Ptgs1,Ggtla1,Fads2,Fads1,Prkag1,
organic acid biosynthesis	2.46	0.0215	0.095	Alox12e,Scd1,Ptgs1,Ggtla1,Fads2,Fads1,Prkag1,
nucleotide receptor activity, G-protein coupled	2.309	0.0223	0.097	Ffar2,Sstr2,Sstr4,Gpr15,Bdkrb2,Cxcr4,Cxcr3,Adora2b,

Annotation	Enrichment ratio	p	FDR	Gene symbols
purinergic nucleotide receptor activity, G-protein coupled	2.309	0.0223	0.097	Ffar2,Sstr2,Sstr4,Gpr15,Bdkrb2,Cxcr4,Cxcr3,Adora2b,
purinergic nucleotide receptor activity	2.309	0.0223	0.097	Ffar2,Sstr2,Sstr4,Gpr15,Bdkrb2,Cxcr4,Cxcr3,Adora2b,
nucleotide receptor activity	2.309	0.0223	0.097	Ffar2,Sstr2,Sstr4,Gpr15,Bdkrb2,Cxcr4,Cxcr3,Adora2b,
eye development	2.621	0.0226	0.098	Lim2,Cryab,Crygb,Mitf,Rpgrip1,Pax6,
lymphocyte differentiation	2.621	0.0226	0.098	Il7,Sfpi1,Zap70,Gadd45g,Hmgb3,Nfam1,
cell migration	1.837	0.0227	0.097	Tie1,Efnb1,Kdr,Csf3r,Cck,Robo4,Il17b,Chl1,Itgb3,Cdh2,Lmx1b,Cxcr4,Sst,Pax6,Lama1,
positive regulation of T cell activation	3.233	0.0232	0.098	Il7,Efnb1,Prkcq,Zap70,
eye development (sensu Mammalia)	3.233	0.0232	0.098	Cryab,Crygb,Mitf,Pax6,

Table 3.3- Gene ontology of downregulated genes in Oct4λVP2 cell line.

Gene ontology (GO) annotations of downregulated genes in Oct4λVP2 cell line with more than 2 fold change performed by A.Sharov (NIH)

Annotation	Enrichment ratio	p	FDR	Gene symbols
transcription factor activity	2.27	0	0	Rhox11,Egr1,Cart1,Hod,Trpv6,Asb9,Foxd1,Foxc1,Sox4,Egr2,Wt1,Stat4,Esrrg,Gata4,Meis1,Tcfec,Gata6,Grhl2,Trpc3,Pou5f1,Gata3,Sox17,Hmx3,Msx1,Foxg1,Jun,Ovol1,Isl2,Hlf,Mpdz,Ripk4,Rxrg,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Nkx6-3,Tbx21,Bach2,Foxp1,Rhox10,Foxf2,Tsc22d3,Nkx3-1,Dmrt2,Cebpb,Ar,Irx3,Onecut1,Dlx4,Hoxa7,Mkx,Asb4,Emx1,Runx1t1,Satb1,Hoxb4,Hand1,Dmrt3,Foxq1,Irx2,Spib,Gata5,Taf13,Nr4a2,Ipf1,Lef1,Foxl2,Foxc2,Rai14,Irf8,Dmrtb1,Irx5,Jund1,Nr2c1,Pou3f3,Hes6,Barx1,Trim30,En2,Invs,Gbx2,Nr2f1,Apol6,Elf1,Lhx2,Prox1,Tbx6,
transcription regulator activity	1.964	0	0	Rhox11,Egr1,Cart1,Hod,Trpv6,Asb9,Foxd1,Foxc1,Id1,Sox4,Egr2,Wt1,Stat4,Esrrg,Gata4,Meis1,Tcfec,Gata6,Grhl2,Trpc3,Zfpm2,Pou5f1,Gata3,Sox17,Hmx3,Msx1,Foxg1,Jun,Ovol1,Isl2,Hlf,Lyl1,Id2,Mpdz,Ripk4,Rxrg,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Nkx6-3,Tbx21,Bach2,Nab1,Foxp1,Rhox10,Bcl6b,Foxf2,Ncoa2,Tsc22d3,Nkx3-1,Bcl11a,Dmrt2,Cebpb,Ar,Irx3,Onecut1,Dlx4,Hoxa7,Mkx,Asb4,Emx1,Runx1t1,Satb1,Hoxb4,Hand1,Dmrt3,Foxq1,Irx2,Spib,Gata5,Taf13,Nr4a2,Ipf1,Lef1,Foxl2,Foxc2,Rai14,Irf8,Dmrtb1,Irx5,Jund1,Nr2c1,Pou3f3,Hes6,Barx1,Ptrf,Trim30,Ascl2,Rest,En2,Invs,Gbx2,Nr2f1,Apol6,Rnf6,Elf1,Wwtr1,Lhx2,Prox1,Gtf2a1lf,Tbx6,
regulation of cellular process	1.571	0	0	Rhox11,Egr1,Cart1,Hod,Trpv6,Asb9,Foxd1,Il10,Foxc1,Egr4,Id1,Msl31,Sox4,Igfbp3,Xdh,Egr2,Rnf128,Wt1,Lta,Stat4,Esrrg,Gata4,Meis1,Tcfec,Ptx3,Gata6,Grhl2,Fst,Trpc3,Zfpm2,Pou5f1,Gata3,Sox17,Plcb1,Hrasls,Hmx3,Msx1,Foxg1,Jun,Ovol1,Lmo4,Fgf5,Isl2,Hlf,Lyl1,Id2,Mpdz,Fosl1,Ripk4,Rxrg,Pogk,C1qtnf2,Cpeb1,Sfrp4,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Ifnz,Nkx6-3,Tbx21,Bach2,Nab1,Foxp1,Rfxdc1,Rhox10,Fhl2,Map3k8,Reln,Bcl6b,Foxf2,Tgfa,Ncoa2,Chrdl2,Tsc22d3,Nkx3-1,Bcl11a,Dmrt2,Cebpb,Ar,Il3,Irx3,Shh,Gsn,Zf,Onecut1,Amot,Il1b,Dap,Dlx4,Bnc1,Rhob,Prdm8,Bmper,Hoxa7,Mkx,Cebpd,Asb4,Emx1,Runx1t1,Satb1,Zfp94,Egf,Hoxb4,Hand1,Pmaip1,Kcnh8,Dmrt3,Tnfsf13b,Bcl2l1,Foxq1,Zfp422,Scml2,Irx2,Kras,Lama4,Cyr61,Kiss1r,Nog,Zfp52,Spib,Klf3,Gata5,Dkk3,Prkca,Rab5c,Taf13,Zfp641,Nr4a2,Ipf1,Shprh,Cables1,Pak7,Cebpg,Lef1,Foxl2,Foxc2,Rai14,Prkce,Sirt4,Irf8,Klf7,Spred2,Dmrtb1,Irx5,Jund1,Rab15,Nr2c1,Pou3f3,Hes6,Barx1,Ptrf,Cebpa,Cnksr2,Klf6,Trim30,Spry1,Ascl2,Gas2,Kcnh5,Rest,En2,Invs,Gbx2,Zfp54,Nr2f1,Apol6,Rnf6,Elf1,Dock2,Per2,Fgf8,Igf2,Wwtr1,Lhx2,Prox1,Camk1d,Rem2,Perp,Phf19,Gtf2a1lf,Rps4x,Nsbp1,Setbp1,Marcks,Tbx6,

Annotation	Enrichment ratio	p	FDR	Gene symbols
regulation of metabolism	1.677	0	0	Rhox11,Egr1,Cart1,Hod,Trpv6,Asb9,Foxd1,Il10,Foxc1,Egr4,Id1,Msl31,Sox4,Egr2,Rnf128,Wt1,Stat4,Esrrg,Gata4,Meis1,Tcfec,Ptx3,Gata6,Grhl2,Trpc3,Zfpm2,Pou5f1,Gata3,Sox17,Hmx3,Msx1,Foxg1,Jun,Ovol1,Lmo4,Isl2,Hlf,Lyl1,Id2,Mpdz,Fosl1,Ripk4,Rxrg,Pogk,Cpeb1,Sfrp4,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Nkx6-3,Tbx21,Bach2,Nab1,Foxp1,Rfxdc1,Rhox10,Fhl2,Bcl6b,Foxf2,Ncoa2,Tsc22d3,Nkx3-1,Bcl11a,Dmrt2,Cebpb,Ar,Il3,Irx3,Shh,Gsn,Zf,Onecut1,Il1b,Dlx4,Bnc1,Prdm8,Hoxa7,Mkx,Cebpd,Asb4,Emx1,Runx1t1,Satb1,Zfp94,Egf,Hoxb4,Hand1,Kcnh8,Dmrt3,Foxq1,Zfp422,Scml2,Irx2,Zfp52,Spib,Klf3,Gata5,Prkca,Taf13,Zfp641,Nr4a2,Ipf1,Shprh,Cebpg,Lef1,Foxl2,Foxc2,Rai14,Prkce,Sirt4,Irf8,Klf7,Dmrtb1,Irx5,Jund1,Rab15,Nr2c1,Pou3f3,Hes6,Barx1,Ptrf,Cebpa,Klf6,Trim30,Ascl2,Kcnh5,Rest,En2,Invs,Gbx2,Zfp54,Nr2f1,Apol6,Rnf6,Elf1,Per2,Wwtr1,Lhx2,Prox1,Rem2,Phf19,Gtf2a1lf,Nsbp1,Setbp1,Marcks,Tbx6,
regulation of cellular metabolism	1.679	0	0	Rhox11,Egr1,Cart1,Hod,Trpv6,Asb9,Foxd1,Il10,Foxc1,Egr4,Id1,Msl31,Sox4,Egr2,Rnf128,Wt1,Stat4,Esrrg,Gata4,Meis1,Tcfec,Gata6,Grhl2,Trpc3,Zfpm2,Pou5f1,Gata3,Sox17,Hmx3,Msx1,Foxg1,Jun,Ovol1,Lmo4,Isl2,Hlf,Lyl1,Id2,Mpdz,Fosl1,Ripk4,Rxrg,Pogk,Cpeb1,Sfrp4,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Nkx6-3,Tbx21,Bach2,Nab1,Foxp1,Rfxdc1,Rhox10,Fhl2,Bcl6b,Foxf2,Ncoa2,Tsc22d3,Nkx3-1,Bcl11a,Dmrt2,Cebpb,Ar,Il3,Irx3,Zf,Onecut1,Il1b,Dlx4,Bnc1,Prdm8,Hoxa7,Mkx,Cebpd,Asb4,Emx1,Runx1t1,Satb1,Zfp94,Egf,Hoxb4,Hand1,Kcnh8,Dmrt3,Foxq1,Zfp422,Scml2,Irx2,Zfp52,Spib,Klf3,Gata5,Prkca,Taf13,Zfp641,Nr4a2,Ipf1,Shprh,Cebpg,Lef1,Foxl2,Foxc2,Rai14,Prkce,Sirt4,Irf8,Klf7,Dmrtb1,Irx5,Jund1,Rab15,Nr2c1,Pou3f3,Hes6,Barx1,Ptrf,Cebpa,Klf6,Trim30,Ascl2,Kcnh5,Rest,En2,Invs,Gbx2,Zfp54,Nr2f1,Apol6,Rnf6,Elf1,Per2,Wwtr1,Lhx2,Prox1,Rem2,Phf19,Gtf2a1lf,Nsbp1,Setbp1,Marcks,Tbx6,
regulation of transcription, DNA-dependent	1.719	0	0	Rhox11,Egr1,Cart1,Hod,Trpv6,Asb9,Foxd1,Foxc1,Egr4,Id1,Msl31,Sox4,Egr2,Wt1,Stat4,Esrrg,Gata4,Meis1,Tcfec,Gata6,Trpc3,Zfpm2,Pou5f1,Gata3,Sox17,Hmx3,Msx1,Foxg1,Jun,Ovol1,Lmo4,Isl2,Hlf,Lyl1,Mpdz,Fosl1,Ripk4,Rxrg,Pogk,Sfrp4,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Nkx6-3,Tbx21,Bach2,Nab1,Foxp1,Rfxdc1,Rhox10,Fhl2,Bcl6b,Foxf2,Ncoa2,Tsc22d3,Nkx3-1,Bcl11a,Dmrt2,Cebpb,Ar,Irx3,Zf,Onecut1,Dlx4,Bnc1,Prdm8,Hoxa7,Mkx,Cebpd,Asb4,Emx1,Runx1t1,Satb1,Zfp94,Hoxb4,Hand1,Kcnh8,Dmrt3,Foxq1,Zfp422,Irx2,Zfp52,Spib,Klf3,Gata5,Taf13,Zfp641,Nr4a2,Ipf1,Shprh,Cebpg,Lef1,Foxl2,Foxc2,Rai14,Sirt4,Irf8,Klf7,Dmrtb1,Irx5,Jund1,Rab15,Nr2c1,Pou3f3,Hes6,Barx1,Ptrf,Cebpa,Klf6,Trim30,Kcnh5,Rest,En2,Invs,Gbx2,Zfp54,Nr2f1,Apol6,Rnf6,Elf1,Per2,Wwtr1,Lhx2,Prox1,Rem2,Phf19,Gtf2a1lf,Nsbp1,Setbp1,Marcks,Tbx6,
regulation of transcription	1.7	0	0	Rhox11,Egr1,Cart1,Hod,Trpv6,Asb9,Foxd1,Foxc1,Egr4,Id1,Msl31,Sox4,Egr2,Wt1,Stat4,Esrrg,Gata4,Meis1,Tcfec,Gata6,Grhl2,Trpc3,Zfpm2,Pou5f1,Gata3,Sox17,Hmx3,Msx1,Foxg1,Jun,Ovol1,Lmo4,Isl2,Hlf,Lyl1,Id2,Mpdz,Fosl1,Ripk4,Rxrg,Pogk,Sfrp4,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Nkx6-3,Tbx21,Bach2,Nab1,Foxp1,Rfxdc1,Rhox10,Fhl2,Bcl6b,Foxf2,Ncoa2,Tsc22d3,Nkx3-1,Bcl11a,Dmrt2,Cebpb,Ar,Irx3,Zf,Onecut1,Dlx4,Bnc1,Prdm8,Hoxa7,Mkx,Cebpd,Asb4,Emx1,Runx1t1,Satb1,Zfp94,Hoxb4,Hand1,Kcnh8,Dmrt3,Foxq1,Zfp422,Scml2,Irx2,Zfp52,Spib,Klf3,Gata5,Taf13,Zfp641,Nr4a2,Ipf1,Shprh,Cebpg,Lef1,Foxl2,Foxc2,Rai14,Sirt4,Irf8,Klf7,Dmrtb1,Irx5,Jund1,Rab15,Nr2c1,Pou3f3,Hes6,Barx1,Ptrf,Cebpa,Klf6,Trim30,Ascl2,Kcnh5,Rest,En2,Invs,Gbx2,Zfp54,Nr2f1,Apol6,Rnf6,Elf1,Per2,Wwtr1,Lhx2,Prox1,Rem2,Phf19,Gtf2a1lf,Nsbp1,Setbp1,Marcks,Tbx6,

Annotation	Enrichment ratio	P	FDR	Gene symbols
development	1.732	0	0	Mcf2,Sp8,Egr1,Cart1,Hod,Aph1a,Sema6a,Ephb1,Foxd1,Flt1,Foxc1,Id1,Igfbp3,Xdh,Egr2,Wt1,Zic1,Gata4,Tcfec,Chrdl1,Artn,Gata6,Met,Wnt5a,Fst,Efnb2,Gata3,Hrasls,Hmx3,Msx1,Lamb1-1,Acta1,Slitrk3,Jun,Ovol1,Lmo4,Fgf5,Isl2,Id2,Thbs1,Sema3c,Ank,Pogk,Sfrp4,Thrb,Emx2,Lrp5,Wnt2,Vax2,Ifnz,Tbx21,Wnt4,Jph2,Reln,Angpt1,Tgfa,Chrdl2,Scn8a,Nkx3-1,Smo,Bcl11a,Dmrt2,Crb3,Cebpb,Ar,Wnt9a,Agtr1a,Shh,Onecut1,Amot,Dlx4,Rhob,Hsd11b1,Ifrd1,Mdfi,Hoxa7,Mcoln3,Epb4.1l5,Emx1,Tnfrsf12a,Fgf15,Runx1t1,Egf,Fzd6,Hoxb4,Fzd1,Hand1,Ctf2,Dmrt3,Foxq1,Kras,Sema3e,Lama4,Cyr61,Nog,Plxna3,Ptger4,Dkk3,Wnt7a,Bmp5,Nr4a2,Ipf1,Palm,Cables1,Lef1,Irf8,Spred2,Dmrtb1,Hes6,Spo11,Cebpa,Spryd3,Ptk7,Spry1,Ascl2,Gas2,Htr2b,En2,Dcamkl1,Bmpr1b,Gbx2,Nr2f1,Rnf6,Dock2,Fgf8,Igf2,Lhx2,Prox1,Rps4x,Jph1,Kif5c,Myl1,Tbx6,
regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism	1.685	0	0	Rhox11,Egr1,Cart1,Hod,Trpv6,Asb9,Foxd1,Foxc1,Egr4,Id1,Msl31,Sox4,Egr2,Wt1,Stat4,Esrrg,Gata4,Meis1,Tcfec,Gata6,Grhl2,Trpc3,Zfpm2,Pou5f1,Gata3,Sox17,Hmx3,Msx1,Foxg1,Jun,Ovol1,Lmo4,Isl2,Hlf,Lyl1,Id2,Mpdz,Fosl1,Ripk4,Rxrg,Pogk,Sfrp4,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Nkx6-3,Tbx21,Bach2,Nab1,Foxp1,Rfxdc1,Rhox10,Fhl2,Bcl6b,Foxf2,Ncoa2,Tsc22d3,Nkx3-1,Bcl11a,Dmrt2,Cebpb,Ar,Irx3,Zf,Onecut1,Dlx4,Bnc1,Prdm8,Hoxa7,Mkx,Cebpd,Asb4,Emx1,Runx1t1,Satb1,Zfp94,Hoxb4,Hand1,Kcnh8,Dmrt3,Foxq1,Zfp422,Scml2,Irx2,Zfp52,Spib,Klf3,Gata5,Taf13,Zfp641,Nr4a2,Ipf1,Shprh,Cebpg,Lef1,Foxl2,Foxc2,Rai14,Sirt4,Irf8,Klf7,Dmrtb1,Irx5,Jund1,Rab15,Nr2c1,Pou3f3,Hes6,Barx1,Ptrf,Cebpa,Klf6,Trim30,Ascl2,Kcnh5,Rest,En2,Invs,Gbx2,Zfp54,Nr2f1,Apol6,Rnf6,Elf1,Per2,Wwtr1,Lhx2,Prox1,Rem2,Phf19,Gtf2a1lf,Nsbp1,Setbp1,Marcks,Tbx6,
transcription, DNA-dependent	1.693	0	0	Rhox11,Egr1,Cart1,Hod,Trpv6,Asb9,Foxd1,Foxc1,Egr4,Id1,Msl31,Sox4,Egr2,Wt1,Stat4,Esrrg,Gata4,Meis1,Tcfec,Gata6,Trpc3,Zfpm2,Pou5f1,Gata3,Sox17,Hmx3,Msx1,Foxg1,Jun,Ovol1,Lmo4,Isl2,Hlf,Lyl1,Mpdz,Fosl1,Ripk4,Rxrg,Pogk,Sfrp4,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Nkx6-3,Tbx21,Bach2,Nab1,Foxp1,Rfxdc1,Rhox10,Fhl2,Bcl6b,Foxf2,Ncoa2,Tsc22d3,Nkx3-1,Bcl11a,Dmrt2,Cebpb,Ar,Irx3,Zf,Onecut1,Dlx4,Bnc1,Prdm8,Hoxa7,Mkx,Cebpd,Asb4,Emx1,Runx1t1,Satb1,Zfp94,Hoxb4,Hand1,Kcnh8,Dmrt3,Foxq1,Zfp422,Irx2,Zfp52,Spib,Klf3,Gata5,Taf13,Zfp641,Nr4a2,Ipf1,Shprh,Cebpg,Lef1,Foxl2,Foxc2,Rai14,Sirt4,Irf8,Klf7,Dmrtb1,Irx5,Jund1,Rab15,Nr2c1,Pou3f3,Hes6,Barx1,Ptrf,Cebpa,Klf6,Trim30,Kcnh5,Rest,En2,Invs,Gbx2,Zfp54,Nr2f1,Apol6,Rnf6,Elf1,Per2,Wwtr1,Lhx2,Prox1,Rem2,Phf19,Gtf2a1lf,Nsbp1,Setbp1,Marcks,Tbx6,

Annotation	Enrichment ratio	p	FDR	Gene symbols
regulation of biological process	1.515	0	0	Rhox11,Egr1,Cart1,Hod,Aph1a,Trpv6,Asb9,Foxd1,Il10,Foxc1,Egr4,Id1,Msl31,Sox4,Igfbp3,Pkia,Xdh,Egr2,Rnf128,Wt1,Lta,Stat4,Esrrg,Gata4,Meis1,Tcfec,Ptx3,Gata6,Grhl2,Fst,Trpc3,Zfpm2,Pou5f1,Gata3,Sox17,Plcb1,Hrasls,Hmx3,Msx1,Foxg1,Jun,Ovol1,Lmo4,Fgf5,Isl2,Hlf,Lyl1,Id2,Mpdz,Fosl1,Ank,Ripk4,Rxrg,Pogk,C1qtnf2,Cpeb1,Sfrp4,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Ifnz,Nkx6-3,Tbx21,Bach2,Nab1,Foxp1,Rfxdc1,Rhox10,Fhl2,Map3k8,Reln,Bcl6b,Foxf2,Tgfa,Ncoa2,Anxa8,Chrdl2,Tsc22d3,Nkx3-1,Bcl11a,Dmrt2,Cebpb,Ar,Il3,Irx3,Shh,Gsn,Zf,Onecut1,Amot,Gnao1,Il1b,Dap,Dlx4,Bnc1,Rhob,Prdm8,Bmper,Hoxa7,Mkx,Cebpd,Asb4,Emx1,Tnfrsf12a,Runx1t1,Satb1,Zfp94,Egf,Hoxb4,Hand1,Pmaip1,Kcnh8,Dmrt3,Tnfsf13b,Bcl2l1,Anxa10,Foxq1,Zfp422,Scml2,Irx2,Kras,Lama4,Cyr61,Kiss1r,Nog,Zfp52,Tshr,Spib,Klf3,Ptger4,Gata5,Dkk3,Prkca,Rab5c,Taf13,Zfp641,Nr4a2,Ipf1,Shprh,Cables1,Pak7,Cebpg,Lef1,Foxl2,Foxc2,Rai14,Prkce,Sirt4,Irf8,Klf7,Spred2,Dmrtb1,Irx5,Jund1,Rab15,Nr2c1,Pou3f3,Hes6,Barx1,Ptrf,Cebpa,Cnksr2,Klf6,Trim30,Spry1,Ascl2,Gas2,Kcnh5,Rest,En2,Invs,Gbx2,Zfp54,Nr2f1,Apol6,Rnf6,Elf1,Dock2,Per2,Fgf8,Igf2,Wwtr1,Lhx2,Prox1,Camk1d,Rem2,Perp,Phf19,Gtf2a1lf,Rps4x,Nsbp1,Setbp1,Marcks,Tbx6,
regulation of physiological process	1.535	0	0	Rhox11,Egr1,Cart1,Hod,Trpv6,Asb9,Foxd1,Il10,Foxc1,Egr4,Id1,Msl31,Sox4,Igfbp3,Egr2,Rnf128,Wt1,Lta,Stat4,Esrrg,Gata4,Meis1,Tcfec,Ptx3,Gata6,Grhl2,Trpc3,Zfpm2,Pou5f1,Gata3,Sox17,Plcb1,Hrasls,Hmx3,Msx1,Foxg1,Jun,Ovol1,Lmo4,Fgf5,Isl2,Hlf,Lyl1,Id2,Mpdz,Fosl1,Ank,Ripk4,Rxrg,Pogk,C1qtnf2,Cpeb1,Sfrp4,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Ifnz,Nkx6-3,Tbx21,Bach2,Nab1,Foxp1,Rfxdc1,Rhox10,Fhl2,Map3k8,Bcl6b,Foxf2,Tgfa,Ncoa2,Anxa8,Tsc22d3,Nkx3-1,Bcl11a,Dmrt2,Cebpb,Ar,Il3,Irx3,Shh,Gsn,Zf,Onecut1,Amot,Gnao1,Il1b,Dap,Dlx4,Bnc1,Rhob,Prdm8,Hoxa7,Mkx,Cebpd,Asb4,Emx1,Runx1t1,Satb1,Zfp94,Egf,Hoxb4,Hand1,Pmaip1,Kcnh8,Dmrt3,Tnfsf13b,Bcl2l1,Anxa10,Foxq1,Zfp422,Scml2,Irx2,Kras,Lama4,Cyr61,Kiss1r,Zfp52,Tshr,Spib,Klf3,Ptger4,Gata5,Prkca,Rab5c,Taf13,Zfp641,Nr4a2,Ipf1,Shprh,Cables1,Pak7,Cebpg,Lef1,Foxl2,Foxc2,Rai14,Prkce,Sirt4,Irf8,Klf7,Dmrtb1,Irx5,Jund1,Rab15,Nr2c1,Pou3f3,Hes6,Barx1,Ptrf,Cebpa,Klf6,Trim30,Ascl2,Gas2,Kcnh5,Rest,En2,Invs,Gbx2,Zfp54,Nr2f1,Apol6,Rnf6,Elf1,Per2,Fgf8,Igf2,Wwtr1,Lhx2,Prox1,Camk1d,Rem2,Perp,Phf19,Gtf2a1lf,Rps4x,Nsbp1,Setbp1,Marcks,Tbx6,
regulation of cellular physiological process	1.537	0	0	Rhox11,Egr1,Cart1,Hod,Trpv6,Asb9,Foxd1,Il10,Foxc1,Egr4,Id1,Msl31,Sox4,Igfbp3,Egr2,Rnf128,Wt1,Lta,Stat4,Esrrg,Gata4,Meis1,Tcfec,Ptx3,Gata6,Grhl2,Trpc3,Zfpm2,Pou5f1,Gata3,Sox17,Plcb1,Hrasls,Hmx3,Msx1,Foxg1,Jun,Ovol1,Lmo4,Fgf5,Isl2,Hlf,Lyl1,Id2,Mpdz,Fosl1,Ripk4,Rxrg,Pogk,C1qtnf2,Cpeb1,Sfrp4,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Ifnz,Nkx6-3,Tbx21,Bach2,Nab1,Foxp1,Rfxdc1,Rhox10,Fhl2,Map3k8,Bcl6b,Foxf2,Tgfa,Ncoa2,Tsc22d3,Nkx3-1,Bcl11a,Dmrt2,Cebpb,Ar,Il3,Irx3,Shh,Gsn,Zf,Onecut1,Amot,Il1b,Dap,Dlx4,Bnc1,Rhob,Prdm8,Hoxa7,Mkx,Cebpd,Asb4,Emx1,Runx1t1,Satb1,Zfp94,Egf,Hoxb4,Hand1,Pmaip1,Kcnh8,Dmrt3,Tnfsf13b,Bcl2l1,Foxq1,Zfp422,Scml2,Irx2,Kras,Lama4,Cyr61,Kiss1r,Zfp52,Spib,Klf3,Gata5,Prkca,Rab5c,Taf13,Zfp641,Nr4a2,Ipf1,Shprh,Cables1,Pak7,Cebpg,Lef1,Foxl2,Foxc2,Rai14,Prkce,Sirt4,Irf8,Klf7,Dmrtb1,Irx5,Jund1,Rab15,Nr2c1,Pou3f3,Hes6,Barx1,Ptrf,Cebpa,Klf6,Trim30,Ascl2,Gas2,Kcnh5,Rest,En2,Invs,Gbx2,Zfp54,Nr2f1,Apol6,Rnf6,Elf1,Per2,Fgf8,Igf2,Wwtr1,Lhx2,Prox1,Camk1d,Rem2,Perp,Phf19,Gtf2a1lf,Rps4x,Nsbp1,Setbp1,Marcks,Tbx6,

Annotation	Enrichment ratio	p	FDR	Gene symbols
transcription	1.641	0	0	Rhox11,Egr1,Cart1,Hod,Trpv6,Asb9,Foxd1,Foxc1,Egr4,Id1,Msl31,Sox4,Egr2,Wt1,Stat4,Esrrg,Gata4,Meis1,Tcfec,Gata6,Grhl2,Trpc3,Zfpm2,Pou5f1,Gata3,Sox17,Hmx3,Msx1,Foxg1,Jun,Ovol1,Lmo4,Isl2,Hlf,Lyl1,Id2,Mpdz,Fosl1,Ripk4,Rxrg,Pogk,Sfrp4,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Nkx6-3,Tbx21,Bach2,Nab1,Foxp1,Rfxdc1,Rhox10,Fhl2,Bcl6b,Foxf2,Ncoa2,Tsc22d3,Nkx3-1,Bcl11a,Dmrt2,Cebpb,Ar,Irx3,Zf,Onecut1,Dlx4,Bnc1,Prdm8,Hoxa7,Mkx,Cebpd,Asb4,Emx1,Runx1t1,Satb1,Zfp94,Hoxb4,Hand1,Kcnh8,Dmrt3,Foxq1,Zfp422,Scml2,Irx2,Zfp52,Spib,Klf3,Gata5,Taf13,Zfp641,Nr4a2,Ipf1,Shprh,Cebpg,Lef1,Foxl2,Foxc2,Rai14,Sirt4,Irf8,Klf7,Dmrtb1,Irx5,Jund1,Rab15,Nr2c1,Pou3f3,Hes6,Barx1,Ptrf,Cebpa,Klf6,Trim30,Ascl2,Kcnh5,Rest,En2,Invs,Gbx2,Zfp54,Nr2f1,Apol6,Rnf6,Elf1,Per2,Wwtr1,Lhx2,Prox1,Rem2,Phf19,Gtf2a1lf,Nsbp1,Setbp1,Marcks,Tbx6,
cell differentiation	2.037	0	0	Mcf2,Egr1,Sema6a,Ephb1,Foxd1,Flt1,Xdh,Egr2,Wt1,Zic1,Artn,Fst,Efnb2,Gata3,Slitrk3,Jun,Fgf5,Sema3c,Sfrp4,Emx2,Ifnz,Reln,Angpt1,Chrdl2,Smo,Bcl11a,Crb3,Cebpb,Shh,Onecut1,Amot,Rhob,Ifrd1,Mdfi,Mcoln3,Emx1,Tnfrsf12a,Fgf15,Runx1t1,Kras,Sema3e,Nog,Bmp5,Nr4a2,Lef1,Irf8,Hes6,Cebpa,Ptk7,Spry1,Ascl2,En2,Dcamkl1,Bmpr1b,Gbx2,Rnf6,Dock2,Fgf8,Prox1,Kif5c,Tbx6,
diacylglycerol metabolism	25.306	0	0	Mogat1,Cds1,Dgkh,
molybdenum ion binding	25.306	0	0	Aox3,Xdh,Aox1,
collagen binding	25.306	0	0	Matn4,Matn3,Mrc2,
organ development	1.877	0	0	Egr1,Hod,Aph1a,Flt1,Foxc1,Id1,Wt1,Gata4,Gata6,Met,Wnt5a,Fst,Efnb2,Gata3,Hmx3,Msx1,Acta1,Id2,Thbs1,Ank,Thrb,Emx2,Wnt2,Ifnz,Wnt4,Reln,Angpt1,Tgfa,Scn8a,Smo,Bcl11a,Cebpb,Ar,Agtr1a,Shh,Amot,Rhob,Hsd11b1,Ifrd1,Hoxa7,Emx1,Tnfrsf12a,Fgf15,Hoxb4,Hand1,Kras,Lama4,Cyr61,Nog,Ptger4,Wnt7a,Bmp5,Lef1,Irf8,Spo11,Cebpa,Spry1,Htr2b,En2,Bmpr1b,Nr2f1,Dock2,Fgf8,Igf2,Lhx2,Prox1,Jph1,Myl1,Tbx6,
DNA binding	1.531	0	0	Rhox11,Ddx3y,Sp8,Egr1,Cart1,Hod,Trpv6,Asb9,Foxd1,Foxc1,Egr4,Sox4,Egr2,Wt1,Zic1,Stat4,Esrrg,Gata4,Meis1,Tcfec,Gata6,Grhl2,Trpc3,Zfpm2,Pou5f1,Gata3,Sox17,Hmx3,Msx1,Foxg1,Jun,Hist3h2ba,Ovol1,Isl2,Hlf,Lyl1,Mpdz,Fosl1,Ripk4,Rxrg,Sfrp4,Thrb,Hlxb9,Emx2,Tbx15,Grm1,Hmx1,Prpf19,Trpc7,Atf5,Pou6f1,Foxa3,Vax2,Nkx6-3,Tbx21,Bach2,Foxp1,Rfxdc1,Rhox10,Bcl6b,Foxf2,Tsc22d3,Nkx3-1,Dmrt2,Cebpb,Ar,Irx3,Recql,Sos2,Zf,Onecut1,Dlx4,Bnc1,Prdm8,Zfp560,Hoxa7,Mkx,Cebpd,Asb4,Emx1,Runx1t1,Satb1,Ppfibp1,Hoxb4,Hand1,Dmrt3,Hnrpul2,Foxq1,Zfp422,Irx2,Spib,Klf3,Hist1h1a,Gata5,Taf13,Nr4a2,Ipf1,Shprh,Bahcc1,Cebpg,Lef1,Foxl2,Foxc2,Rai14,Sirt4,Irf8,Klf7,Dmrtb1,Irx5,Jund1,Rab15,Nr2c1,Pou3f3,Hes6,Barx1,Spo11,Cebpa,Klf6,Trim30,Ascl2,Rest,En2,Invs,Gbx2,Uhmk1,Nr2f1,Apol6,Rnf6,Elf1,Lhx2,Prox1,Rem2,Phf19,Gtf2a1lf,Nsbp1,Rad51l3,Setbp1,Marcks,Tbx6,
morphogenesis	1.905	0	0	Sp8,Hod,Ephb1,Foxd1,Flt1,Id1,Igfbp3,Xdh,Wt1,Gata4,Artn,Gata6,Wnt5a,Fst,Efnb2,Hrasls,Hmx3,Slitrk3,Jun,Lmo4,Id2,Thbs1,Thrb,Lrp5,Wnt2,Wnt4,Reln,Angpt1,Tgfa,Smo,Crb3,Cebpb,Ar,Wnt9a,Agtr1a,Shh,Amot,Rhob,Hsd11b1,Hoxa7,Tnfrsf12a,Fgf15,Egf,Hoxb4,Hand1,Foxq1,Cyr61,Wnt7a,Palm,Lef1,Spo11,Cebpa,Ptk7,Spry1,Gas2,Htr2b,Gbx2,Dock2,Fgf8,Igf2,Kif5c,Tbx6,
system development	2.227	0	0	Mcf2,Aph1a,Sema6a,Ephb1,Foxd1,Egr2,Wt1,Zic1,Artn,Met,Efnb2,Hmx3,Msx1,Slitrk3,Lmo4,Fgf5,Sema3c,Emx2,Reln,Smo,Cebpb,Agtr1a,Shh,Emx1,Tnfrsf12a,Sema3e,Nog,Nr4a2,Cables1,Hes6,Ptk7,Spry1,Ascl2,En2,Dcamkl1,Gbx2,Nr2f1,Fgf8,Lhx2,Kif5c,
neutral lipid metabolism	11.247	0	0	Apob,Mogat1,Cds1,Dgkh,
acylglycerol metabolism	11.247	0	0	Apob,Mogat1,Cds1,Dgkh,

Annotation	Enrichment ratio	p	FDR	Gene symbols
epithelial cell differentiation	8.435	0	0	Xdh,Wt1,Jun,Crb3,Ptk7,
sex differentiation	4.898	0	0	Wt1,Fst,Wnt4,Dmrt2,Ar,Shh,Dmrt3,Dmrtb1,Spo11,
behavior	2.789	0	0	Itga8,Egr2,Zic1,Artn,Met,Cmtm8,Reln,Scn8a,Ccl20,Amot,Gnao1,Il1b,Vdac1,Mcoln3,Cyr61,Tshr,Prkca,Prkce,Cxcl16,Astn1,Dock2,
receptor binding	1.957	0	0	Irs4,Pdyn,Il10,Areg,Irs3,Lta,Artn,Wnt5a,Nrg4,Tnfsf14,Fgf5,Cxcl7,Uts2,Sema3c,Irs1,Wnt2,Ifnz,Pramel3,Wnt4,Cmtm8,Angpt1,Tgfa,Ncoa2,Il3,Calcb,Ccl20,Adm2,Dok2,Il1b,Nsmaf,Fgf15,Egf,Ctf2,Tnfsf13b,Lama4,Fgf21,Fgf13,Wnt7a,Bmp5,Prlpm,Cxcl16,Spred2,Spryd3,Inha,Fgf8,Igf2,Pnoc,
organ morphogenesis	2.101	0	0	Hod,Flt1,Id1,Wt1,Gata4,Gata6,Wnt5a,Fst,Efnb2,Hmx3,Id2,Thbs1,Thrb,Wnt2,Wnt4,Angpt1,Tgfa,Smo,Cebpb,Ar,Shh,Amot,Rhob,Hsd11b1,Hoxa7,Tnfrsf12a,Fgf15,Hoxb4,Hand1,Cyr61,Wnt7a,Lef1,Spo11,Cebpa,Spry1,Htr2b,Fgf8,Igf2,
glycerol ether metabolism	9.64	0	0	Apob,Mogat1,Cds1,Dgkh,
glycerolipid metabolism	9.64	0	0	Apob,Mogat1,Cds1,Dgkh,
appendage morphogenesis	5.368	0	0	Sp8,Wnt5a,Lrp5,Wnt9a,Shh,Wnt7a,Fgf8,
limb morphogenesis	5.368	0	0	Sp8,Wnt5a,Lrp5,Wnt9a,Shh,Wnt7a,Fgf8,
appendage development	5.368	0	0	Sp8,Wnt5a,Lrp5,Wnt9a,Shh,Wnt7a,Fgf8,
nervous system development	2.116	0	0	Mcf2,Sema6a,Ephb1,Foxd1,Egr2,Zic1,Artn,Met,Efnb2,Hmx3,Msx1,Slitrk3,Lmo4,Fgf5,Sema3c,Emx2,Reln,Smo,Cebpb,Shh,Emx1,Tnfrsf12a,Sema3e,Nog,Nr4a2,Cables1,Hes6,Ptk7,Ascl2,En2,Dcamkl1,Gbx2,Nr2f1,Lhx2,Kif5c,
central nervous system development	3.426	0	0	Zic1,Met,Hmx3,Msx1,Emx2,Reln,Smo,Shh,Emx1,Nog,En2,Nr2f1,Lhx2,
establishment of cell polarity	12.653	0	0	Crb3,Ptk7,Dock2,
cell migration	2.656	0	0	Ephb1,Foxd1,Flt1,Artn,Lrp5,Reln,Shh,Amot,Il1b,Tnfrsf12a,Fgf15,Lama4,Kiss1r,Prkca,Srgap1,Gbx2,Nr2f1,Astn1,Dock2,Kif5c,
insulin receptor binding	8.435	0	0	Irs4,Irs3,Irs1,Dok2,
transcription factor complex	1.987	0	0	Foxd1,Foxc1,Sox4,Meis1,Tcfec,Sox17,Hmx3,Msx1,Jun,Lmo4,Hlxb9,Hmx1,Atf5,Pou6f1,Foxa3,Foxf2,Nkx3-1,Irx3,Onecut1,Dlx4,Hoxa7,Hoxb4,Hand1,Foxq1,Irx2,Taf13,Lef1,Foxc2,Pou3f3,Hes6,Barx1,Cebpa,Gbx2,Nr2f1,Wwtr1,Lhx2,Prox1,
vasculature development	3.067	0	0	Flt1,Id1,Efnb2,Thbs1,Angpt1,Tgfa,Smo,Shh,Amot,Rhob,Tnfrsf12a,Hand1,Lama4,Cyr61,
locomotory behavior	2.942	0	0	Artn,Cmtm8,Reln,Scn8a,Ccl20,Amot,Gnao1,Il1b,Mcoln3,Cyr61,Tshr,Prkca,Cxcl16,Astn1,Dock2,
brain development	3.501	0	0	Met,Hmx3,Msx1,Emx2,Reln,Shh,Emx1,Nog,En2,Nr2f1,Lhx2,
embryonic development	2.477	0	0	Sp8,Flt1,Gata4,Wnt5a,Hmx3,Lmo4,Lrp5,Cebpb,Ar,Wnt9a,Shh,Amot,Lama4,Cyr61,Wnt7a,Lef1,Cebpa,Ptk7,Ascl2,Fgf8,Tbx6,
collagen catabolism	7.498	0	0.001	Mmp16,Mmp8,Mmp14,Mmp2,
BMP signaling pathway	6.025	0	0.001	Fst,Chrdl2,Bmper,Nog,Bmpr1b,
lipoprotein binding	10.122	0	0.001	Pon1,Lrp8,Cxcl16,

Annotation	Enrichment ratio	p	FDR	Gene symbols
ion channel activity	1.974	0.0001	0.001	Kctd12b,Trpv6,Kcna1,Cacna2d1,Glrb,Chrnb1,Trpc3,Kcnmb2,Kctd4,Trpc7,Gabra4,Kcnk1,Gabra2,Klhl24,Clic6,Scn8a,Kctd14,Kcne3,Kcnj3,Vdac1,Mcoln3,Kcnmb4,Gabra3,Kcnh8,Kcnk6,Pllp,Scn1a,Cacnb2,Kcnt2,Kcna7,Kcnh5,Glra2,Kctd16,
frizzled-2 signaling pathway	5.623	0.0001	0.001	Wnt5a,Wnt2,Wnt4,Wnt9a,Wnt7a,
blood vessel morphogenesis	3.067	0.0001	0.001	Flt1,Id1,Thbs1,Angpt1,Tgfa,Smo,Shh,Amot,Rhob,Tnfrsf12a,Hand1,Cyr61,
Wnt receptor signaling pathway	2.924	0.0001	0.001	Wnt5a,Sfrp4,Lrp5,Wnt2,Wnt4,Wnt9a,Kremen2,Fzd6,Fzd1,Dkk3,Wnt7a,Lef1,Csnk1e,
alpha-type channel activity	1.91	0.0001	0.002	Kctd12b,Trpv6,Kcna1,Cacna2d1,Glrb,Chrnb1,Trpc3,Kcnmb2,Kctd4,Trpc7,Gabra4,Kcnk1,Gabra2,Klhl24,Clic6,Scn8a,Kctd14,Kcne3,Kcnj3,Vdac1,Mcoln3,Kcnmb4,Gabra3,Kcnh8,Kcnk6,Gja3,Pllp,Scn1a,Cacnb2,Kcnt2,Kcna7,Kcnh5,Glra2,Aqp5,Kctd16,
blood vessel development	2.885	0.0002	0.002	Flt1,Id1,Thbs1,Angpt1,Tgfa,Smo,Shh,Amot,Rhob,Tnfrsf12a,Hand1,Lama4,Cyr61,
embryonic limb morphogenesis	4.601	0.0002	0.002	Sp8,Wnt5a,Lrp5,Wnt9a,Shh,Wnt7a,
embryonic appendage morphogenesis	4.601	0.0002	0.002	Sp8,Wnt5a,Lrp5,Wnt9a,Shh,Wnt7a,
nucleoplasm	1.799	0.0002	0.003	Foxd1,Foxc1,Sox4,Meis1,Tcfec,Sox17,Hmx3,Msx1,Jun,Lmo4,Hlxb9,Hmx1,Atf5,Pou6f1,Foxa3,Foxf2,Nkx3-1,Irx3,Onecut1,Dlx4,Hoxa7,Hoxb4,Hand1,Foxq1,Irx2,Taf13,Lef1,Foxc2,Pou3f3,Hes6,Barx1,Cebpa,Rest,Gbx2,Nr2f1,Rnf6,Wwtr1,Lhx2,Prox1,Nsbp1,
dorsal/ventral pattern formation	5.272	0.0002	0.003	Sp8,Smo,Shh,Bmpr1b,Lhx2,
locomotion	2.256	0.0002	0.003	Ephb1,Foxd1,Flt1,Artn,Lrp5,Reln,Shh,Amot,Il1b,Tnfrsf12a,Fgf15,Lama4,Kiss1r,Tshr,Prkca,Srgap1,Gbx2,Nr2f1,Astn1,Dock2,Kif5c,
fat cell differentiation	8.435	0.0002	0.003	Cebpb,Runx1t1,Cebpa,
sodium:hydrogen antiporter activity	8.435	0.0002	0.002	Slc9a2,Slc9a7,Slc9a5,
monovalent cation:proton antiporter activity	8.435	0.0002	0.002	Slc9a2,Slc9a7,Slc9a5,
channel or pore class transporter activity	1.845	0.0004	0.004	Kctd12b,Trpv6,Kcna1,Cacna2d1,Glrb,Chrnb1,Trpc3,Kcnmb2,Kctd4,Trpc7,Gabra4,Kcnk1,Gabra2,Klhl24,Clic6,Scn8a,Kctd14,Kcne3,Kcnj3,Vdac1,Mcoln3,Kcnmb4,Gabra3,Kcnh8,Kcnk6,Gja3,Pllp,Scn1a,Cacnb2,Kcnt2,Kcna7,Kcnh5,Glra2,Aqp5,Kctd16,
metal ion transport	1.874	0.0005	0.005	Slc39a8,Slc9a2,Kctd12b,Trpv6,Kcna1,Slc23a1,Slc11a1,Cacna2d1,Slc40a1,Trpc3,Atp2a3,Kcnmb2,Slc9a7,Kctd4,Trpc7,Kcnk1,Scn8a,Kctd14,Kcne3,Kcnj3,Kcnmb4,Kcnh8,Kcnk6,Scn1a,Cacnb2,Slc9a5,Kcnt2,Sfxn5,Slc5a5,Kcna7,Kcnh5,Kctd16,
tissue development	2.257	0.0005	0.005	Foxc1,Wt1,Gata3,Ovol1,Ank,Angpt1,Smo,Shh,Onecut1,Mcoln3,Epb4.1l5,Foxq1,Ptger4,Spry1,Bmpr1b,Fgf8,Lhx2,Prox1,Tbx6,
cell motility	2.176	0.0007	0.006	Ephb1,Foxd1,Flt1,Artn,Lrp5,Reln,Shh,Amot,Il1b,Tnfrsf12a,Fgf15,Lama4,Kiss1r,Prkca,Srgap1,Gbx2,Nr2f1,Astn1,Dock2,Kif5c,

Annotation	Enrichment ratio	p	FDR	Gene symbols
localization of cell	2.176	0.0007	0.006	Ephb1,Foxd1,Flt1,Artn,Lrp5,Reln,Shh,Amot,Il1b,Tnfrsf12a,Fgf15,Lama4,Kiss1r,Prkca,Srgap1,Gbx2,Nr2f1,Astn1,Dock2,Kif5c,
morphogenesis of a branching structure	4.686	0.0007	0.006	Flt1,Agtr1a,Shh,Egf,Cyr61,
regulation of angiogenesis	7.23	0.0007	0.007	Id1,Thbs1,Amot,
enzyme linked receptor protein signaling pathway	2.195	0.0008	0.007	Ephb1,Flt1,Epha1,Fst,Angpt1,Chrdl2,Dok2,Bmper,Grb10,Nog,Prkca,Bmp5,Ptprk,Stxbp4,Ptprs,Plat,Bmpr1b,Ptprt,Ptpru,
gonad development	4.439	0.0011	0.01	Wt1,Fst,Wnt4,Ar,Spo11,
development of primary sexual characteristics	4.439	0.0011	0.01	Wt1,Fst,Wnt4,Ar,Spo11,
angiogenesis	2.859	0.0012	0.011	Flt1,Id1,Thbs1,Angpt1,Tgfa,Amot,Rhob,Tnfrsf12a,Hand1,Cyr61,
transmembrane receptor protein tyrosine phosphatase signaling pathway	5.191	0.0012	0.01	Ptprk,Ptprs,Ptprt,Ptpru,
calcium ion binding	1.592	0.0013	0.011	Spna1,Usp9y,Fstl1,Cubn,Pla2g5,Sned1,Atp2a3,Pcdhb17,Thbs1,Pcdhb20,Slc25a24,Matn4,Matn2,Capn2,Cib2,Vwa2,Pcdh17,Ncoa2,Anxa8,Pcdhb21,Pls3,Gsn,Caln1,F13a1,Mmp8,Efemp1,Matn3,Efcab4a,Pcdhb14,Eml2,Egf,Cdh22,Anxa10,Aytl1,Nkd1,Cdh10,Prkca,Lrp8,Hpca,Fkbp10,Syt11,Vldlr,Efhd1,Abtb2,Dsc2,Mmp2,Man1c1,Mucdhl,Myl1,
cell-cell signaling	2.017	0.0013	0.011	Pdyn,Glrb,Chrnb1,Egr2,Wnt5a,Gabra4,Wnt2,Gabra2,Wnt4,Wnt9a,Tph2,Shh,Prima1,Vdac1,Gabra3,Gja3,Wnt7a,Nr4a2,Spry1,Glra2,Fgf8,Pnoc,
urogenital system development	3.749	0.0016	0.012	Aph1a,Wt1,Agtr1a,Shh,Nog,Spry1,
metanephros development	4.217	0.0017	0.013	Aph1a,Wt1,Agtr1a,Nog,Spry1,
non-G-protein coupled 7TM receptor activity	6.326	0.0017	0.013	Smo,Fzd6,Fzd1,
vasculogenesis	6.326	0.0017	0.013	Smo,Shh,Amot,
establishment and/or maintenance of cell polarity	6.326	0.0017	0.013	Crb3,Ptk7,Dock2,
ion transport	1.528	0.0018	0.013	Slc39a8,Slc9a2,Kctd12b,Trpv6,Kcna1,Slc23a1,Slc11a1,Col4a6,Cacna2d1,Glrb,Chrnb1,Slc40a1,Slc16a7,Cthrc1,Trpc3,Atp2a3,C1ql2,Kcnmb2,Slc9a7,Ptger3,Ank,Kctd4,Trpc7,Gabra4,Kcnk1,Gabra2,Clic6,Scn8a,Colec12,Kctd14,Kcne3,Kcnj3,Vdac1,Mcoln3,Kcnmb4,Gabra3,Kcnh8,Kcnk6,Col7a1,Immp2l,Pllp,Scn1a,Cacnb2,Slc9a5,Svop,Kcnt2,Sfxn5,Slco1a5,Slc5a5,Kcna7,Kcnh5,Atp11b,Glra2,Sftpd,Kctd16,
cation channel activity	1.918	0.0019	0.014	Kctd12b,Trpv6,Kcna1,Cacna2d1,Chrnb1,Trpc3,Kcnmb2,Kctd4,Trpc7,Kcnk1,Scn8a,Kctd14,Kcne3,Kcnj3,Mcoln3,Kcnmb4,Kcnh8,Kcnk6,Scn1a,Cacnb2,Kcnt2,Kcna7,Kcnh5,Kctd16,
potassium channel activity	2.239	0.0024	0.018	Kctd12b,Kcna1,Kcnmb2,Kctd4,Kcnk1,Kctd14,Kcne3,Kcnj3,Kcnmb4,Kcnh8,Kcnk6,Kcnt2,Kcna7,Kcnh5,Kctd16,

Annotation	Enrichment ratio	p	FDR	Gene symbols
GABA receptor activity	3.49	0.0029	0.021	Glrb,Gabra4,Gabra2,Gabra3,Gpr156,Glra2,
growth factor activity	2.181	0.0034	0.023	Areg,Artn,Nrg4,Fgf5,Tgfa,Il3,Il1b,Fgf15,Egf,Fgf21,Fgf13,Bmp5,Inha,Fgf8,Igf2,
calcium-activated potassium channel activity	5.623	0.0034	0.023	Kcnmb2,Kcnmb4,Kcnt2,
protein homodimerization activity	3.374	0.0038	0.025	Cebpb,Cebpd,Runx1t1,Hand1,Dgkh,Cebpa,
negative regulation of biological process	1.614	0.0038	0.024	Hod,Id1,Pkia,Rnf128,Wt1,Fst,Zfpm2,Thrb,Atf5,Nab1,Foxp1,Bcl6b,Ncoa2,Anxa8,Chrdl2,Tsc22d3,Cebpb,Il3,Shh,Gsn,Rhob,Bmper,Pmaip1,Tnfsf13b,Bcl2l1,Anxa10,Kiss1r,Nog,Dkk3,Prkca,Pak7,Sirt4,Spred2,Cebpa,Gas2,Rest,Elf1,Prox1,
pattern specification	2.144	0.0041	0.026	Sp8,Flt1,Fst,Emx2,Lrp5,Smo,Shh,Hoxa7,Hoxb4,Cyr61,Nog,Bmp5,Lef1,Bmpr1b,Lhx2,
cytokine activity	1.928	0.0043	0.027	Il10,Areg,Lta,Tnfsf14,Fgf5,Cxcl7,Ifnz,Cmtm8,Angpt1,Il3,Ccl20,Il1b,Fgf15,Ctf2,Tnfsf13b,Fgf13,Bmp5,Cxcl16,Spred2,Fgf8,
axon guidance	3.028	0.0044	0.027	Ephb1,Foxd1,Artn,Shh,Tnfrsf12a,Gbx2,Kif5c,
transcription from RNA polymerase II promoter	1.815	0.0044	0.026	Hod,Foxd1,Id1,Wt1,Gata4,Meis1,Gata6,Zfpm2,Pou5f1,Jun,Lmo4,Hlxb9,Atf5,Fhl2,Bcl6b,Foxf2,Ncoa2,Hand1,Ipf1,Lef1,Foxl2,Hes6,Wwtr1,Gtf2a1lf,
fibroblast growth factor receptor binding	4.217	0.005	0.029	Fgf5,Fgf15,Fgf13,Fgf8,
embryonic morphogenesis	2.617	0.0051	0.029	Sp8,Wnt5a,Hmx3,Lmo4,Lrp5,Wnt9a,Shh,Wnt7a,Ptk7,
small GTPase mediated signal transduction	1.837	0.0055	0.031	Cdc42ep1,Hrasls,Arl4c,Rab27b,Reln,Sos2,Dok2,Rhob,Rnd3,Cdgap,Kras,Rab5c,Rab31,Srgap1,Rab20,Rab15,Arl11,Rgl3,Rhoc,Dock2,Rem2,Mfhas1,
potassium ion transport	2.029	0.006	0.034	Kctd12b,Kcna1,Kcnmb2,Slc9a7,Kctd4,Kcnk1,Kctd14,Kcne3,Kcnj3,Kcnmb4,Kcnh8,Kcnk6,Kcnt2,Kcna7,Kcnh5,Kctd16,
rhythmic process	3.163	0.0062	0.034	Egr2,Hlf,Prokr1,Csnk1e,Spo11,Per2,
negative regulation of signal transduction	3.514	0.0063	0.034	Chrdl2,Bmper,Dkk3,Prkca,Elf1,
kidney development	3.514	0.0063	0.034	Aph1a,Wt1,Agtr1a,Nog,Spry1,
steroid hormone receptor activity	2.88	0.0066	0.035	Esrrg,Rxrg,Thrb,Ar,Nr4a2,Nr2c1,Nr2f1,
ligand-dependent nuclear receptor activity	2.88	0.0066	0.035	Esrrg,Rxrg,Thrb,Ar,Nr4a2,Nr2c1,Nr2f1,
mesoderm development	3.969	0.0072	0.038	Ovol1,Epb4.1l5,Lhx2,Tbx6,
embryonic development (sensu Mammalia)	3.969	0.0072	0.038	Gata4,Ar,Amot,Ascl2,
regulation of peptidyl-tyrosine phosphorylation	3.969	0.0072	0.037	Il3,Egf,Prkca,Prkce,

Annotation	Enrichment ratio	p	FDR	Gene symbols
acetylgalactosaminyltransferase activity	3.969	0.0072	0.036	Galnt7,Galnt14,B3galnt2,Galntl4,
nuclear lumen	1.53	0.0074	0.037	Foxd1,Foxc1,Sox4,Meis1,Tcfec,Sox17,Hmx3,Msx1,Jun,Lmo4,Hlxb9,Hmx1,Atf5,Pou6f1,Foxa3,Snrpd3,Foxf2,Nkx3-1,Irx3,Onecut1,Dlx4,Hoxa7,Hoxb4,Hand1,Foxq1,Irx2,Taf13,Lef1,Foxc2,Pou3f3,Hes6,Barx1,Cebpa,Rest,Gbx2,Nr2f1,Rnf6,Wwtr1,Lhx2,Prox1,Nsbp1,
sulfotransferase activity	3.067	0.0077	0.038	Chst11,Chst1,Hs3st3b1,Chst7,Hs3st1,Sult1b1,
voltage-gated ion channel activity	1.984	0.0078	0.039	Kctd12b,Kcna1,Cacna2d1,Kctd4,Kcnk1,Clic6,Scn8a,Kctd14,Kcne3,Kcnj3,Kcnh8,Scn1a,Cacnb2,Kcna7,Kcnh5,Kctd16,
neurotransmitter receptor activity	2.2	0.0081	0.039	Glrb,Chrnb1,Oprk1,Prokr1,Htr1b,Gabra4,Gabra2,Gabra3,Sort1,Kiss1r,Tacr2,Glra2,
neurotransmitter binding	2.2	0.0081	0.039	Glrb,Chrnb1,Oprk1,Prokr1,Htr1b,Gabra4,Gabra2,Gabra3,Sort1,Kiss1r,Tacr2,Glra2,
neurogenesis	1.924	0.0088	0.043	Mcf2,Ephb1,Foxd1,Egr2,Artn,Slitrk3,Emx2,Reln,Cebpb,Shh,Emx1,Tnfrsf12a,Nr4a2,En2,Gbx2,Nr2f1,Kif5c,
negative regulation of cellular process	1.58	0.009	0.044	Hod,Id1,Rnf128,Wt1,Fst,Zfpm2,Thrb,Atf5,Nab1,Foxp1,Bcl6b,Ncoa2,Chrdl2,Tsc22d3,Cebpb,Il3,Shh,Gsn,Rhob,Bmper,Pmaip1,Tnfsf13b,Bcl2l1,Kiss1r,Nog,Dkk3,Prkca,Pak7,Sirt4,Cebpa,Gas2,Rest,Elf1,Prox1,
regulation of transcription from RNA polymerase II promoter	1.814	0.0097	0.047	Hod,Foxd1,Id1,Wt1,Gata4,Meis1,Gata6,Zfpm2,Jun,Hlxb9,Atf5,Fhl2,Bcl6b,Ncoa2,Hand1,Ipf1,Lef1,Foxl2,Hes6,Wwtr1,
immune cell chemotaxis	4.601	0.0099	0.047	Il1b,Prkca,Dock2,
branching morphogenesis of a tube	4.601	0.0099	0.047	Flt1,Agtr1a,Cyr61,
male sex differentiation	4.601	0.0099	0.047	Wt1,Ar,Shh,
ureteric bud development	4.601	0.0099	0.047	Agtr1a,Nog,Spry1,
regulation of signal transduction	2.279	0.0116	0.054	Reln,Chrdl2,Bmper,Dkk3,Prkca,Spred2,Cnksr2,Spry1,Elf1,Dock2,
neuron differentiation	1.946	0.0124	0.058	Mcf2,Ephb1,Foxd1,Egr2,Artn,Slitrk3,Emx2,Cebpb,Shh,Emx1,Tnfrsf12a,Nr4a2,En2,Gbx2,Kif5c,
GABA-A receptor activity	3.124	0.0134	0.063	Glrb,Gabra4,Gabra2,Gabra3,Glra2,
cell-cell adhesion	2.087	0.0135	0.062	Pcdhb17,Pcdhb20,Pcdh17,Pcdhb21,Scarf2,Pcdhb14,Cdh22,Arvcf,Cdh10,Dsc2,Astn1,Mucdhl,
taxis	2.335	0.0137	0.063	Artn,Cmtm8,Ccl20,Amot,Il1b,Cyr61,Prkca,Cxcl16,Dock2,
chemotaxis	2.335	0.0137	0.063	Artn,Cmtm8,Ccl20,Amot,Il1b,Cyr61,Prkca,Cxcl16,Dock2,
solute:cation antiporter activity	4.217	0.0151	0.07	Slc9a2,Slc9a7,Slc9a5,
solute:hydrogen antiporter activity	4.217	0.0151	0.07	Slc9a2,Slc9a7,Slc9a5,
neurotransmitter metabolism	4.217	0.0151	0.069	Tph2,Prima1,Nr4a2,
adult behavior	4.217	0.0151	0.068	Met,Scn8a,Tshr,

Annotation	Enrichment ratio	p	FDR	Gene symbols
immune cell migration	4.217	0.0151	0.068	Il1b,Prkca,Dock2,
polypeptide N-acetylgalactosaminyltransferase activity	4.217	0.0151	0.068	Galnt7,Galnt14,Galntl4,
extracellular matrix (sensu Metazoa)	1.748	0.0152	0.067	Fras1,Adamts19,Col4a6,Cthrc1,Lamb1-1,Timp3,Matn4,Matn2,Mmp16,Reln,Mmp8,Hc,Matn3,Col7a1,Lama4,Hapln4,Mmp14,Mmp2,Gpc5,Mmp25,
identical protein binding	2.567	0.0157	0.069	Vwa2,Cebpb,Cebpd,Runx1t1,Hand1,Dgkh,Cebpa,
anti-apoptosis	2.567	0.0157	0.068	Atf5,Tsc22d3,Cebpb,Il3,Tnfsf13b,Bcl2l1,Pak7,
tube morphogenesis	3.012	0.0168	0.072	Flt1,Lmo4,Agtr1a,Cyr61,Ptk7,
tight junction	2.663	0.0201	0.082	Cldn7,Cldn1,Cgn,Cldn18,Mpdz,Ocln,
morphogenesis of an epithelium	2.663	0.0201	0.081	Xdh,Wt1,Jun,Lmo4,Crb3,Ptk7,
solute:solute antiporter activity	3.893	0.022	0.088	Slc9a2,Slc9a7,Slc9a5,
circadian rhythm	3.893	0.022	0.087	Prokr1,Csnk1e,Per2,
cation:cation antiporter activity	3.893	0.022	0.087	Slc9a2,Slc9a7,Slc9a5,
negative regulation of physiological process	1.529	0.023	0.092	Hod,Id1,Rnf128,Wt1,Zfpm2,Thrb,Atf5,Nab1,Foxp1,Bcl6b,Ncoa2,Anxa8,Tsc22d3,Cebpb,Il3,Shh,Gsn,Rhob,Pmaip1,Tnfsf13b,Bcl2l1,Anxa10,Kiss1r,Prkca,Pak7,Sirt4,Cebpa,Gas2,Rest,Prox1,
regulation of protein amino acid phosphorylation	3.213	0.0235	0.093	Il3,Egf,Prkca,Prkce,
regulation of amino acid metabolism	3.213	0.0235	0.093	Il3,Egf,Prkca,Prkce,
embryonic development (sensu Vertebrata)	2.811	0.0251	0.1	Gata4,Ar,Amot,Lef1,Ascl2,

Table 3.4- List of the strongest clusters of upregulated genes in Oct4λVP2 cell line.

The strongest 15 clusters of overexpressed genes in Oct4λVP2 cell line relative to Oct4 cell line ($p < 1.0E-6$). The scoring of the clusters integrates information on average overexpression of genes and their proximity.

Chr	Strand	Start	Symbol	Gene name	Log ratio Oct4λVP2/ Oct4	Score
Cluster 1						
chr2	+	25002766	Nrarp	Mus musculus Notch-regulated ankyrin repeat protein (Nrarp), mRNA	0.3137	0.0502
chr2	-	25113189	Grin1	Mus musculus glutamate receptor, ionotropic, NMDA1 (zeta 1) (Grin1), mRNA	1.0629	0.0805
chr2	+	25217882	Entpd2	Mus musculus ectonucleoside triphosphate diphosphohydrolase 2 (Entpd2), mRNA	0.319	0.0791
chr2	+	25278846	Clic3	Mus musculus chloride intracellular channel 3 (Clic3), mRNA	0.4876	0.3902
chr2	-	25397424	Gm996	Mus musculus gene model 996, (NCBI) (Gm996), mRNA	0.3128	1.2649
chr2	-	26174320	Card9	PREDICTED: Mus musculus caspase recruitment domain family, member 9, transcript variant 3 (Card9), mRNA	0.4041	0.4875
chr2	-	26664006	Abo	Mus musculus ABO blood group (transferase A, alpha 1-3-N-acetylgalactosaminyltransferase, transferase B, alpha 1-3-galactosyltransferase) (Abo), mRNA	0.6881	0.3059
chr2	+	27708061	Col5a1	Mus musculus procollagen, type V, alpha 1 (Col5a1), mRNA	0.318	0.182
Cluster 2						
chr2	-	146868986	Nkx2-2	Mus musculus NK2 transcription factor related, locus 2 (Drosophila) (Nkx2-2), mRNA	0.6869	0.0422
chr2	+	148086782	Sstr4	Mus musculus somatostatin receptor 4 (Sstr4), mRNA	0.5515	0.0417
chr2	-	148095906	Thbd	Mus musculus thrombomodulin (Thbd), mRNA	0.5359	0.1218

Chr	Strand	Start	Symbol	Gene name	Log ratio Oct4λVP2/ Oct4	Score
chr2	+	148511521	Cst13	Mus musculus cystatin 13 (Cst13), mRNA	0.7688	0.1356
chr2	+	148526587	Cst9	Mus musculus cystatin 9 (Cst9), mRNA	1.0253	0.4305
chr2	-	148563158	Cst3	Mus musculus cystatin C (Cst3), mRNA	0.3257	1.3465
Cluster 3						
chr5	-	120990949	Oas2	Mus musculus 2'-5' oligoadenylate synthetase 2 (Oas2), mRNA	0.6548	0.0519
chr5	-	121046842	Oas1e	Mus musculus 2'-5' oligoadenylate synthetase 1E (Oas1e), mRNA	0.5155	0.0838
chr5	-	121060709	Oas1c	Mus musculus 2'-5' oligoadenylate synthetase 1C (Oas1c), mRNA	0.536	0.178
chr5	+	121073251	Oas1b	Mus musculus 2'-5' oligoadenylate synthetase 1B (Oas1b), mRNA	0.8709	0.2261
chr5	+	121122037	Oas1h	Mus musculus 2'-5' oligoadenylate synthetase 1H (Oas1h), mRNA	0.4109	4.5589
chr5	+	121175152	Oas1d	Mus musculus 2'-5' oligoadenylate synthetase 1D (Oas1d), mRNA	1.073	5.3095
Cluster 4						
chr6	-	113331801	Tada3l	Mus musculus transcriptional adaptor 3 (NGG1 homolog, yeast)-like (Tada3l), mRNA	0.3642	0.0195
chr6	+	113355036	Ttll3	Mus musculus tubulin tyrosine ligase-like family, member 3 (Ttll3), mRNA	0.3443	0.0223
chr6	+	113437203	Il17rc	Mus musculus interleukin 17 receptor C (Il17rc), mRNA	0.3196	0.0244
chr6	-	113459415	Prrt3	Mus musculus proline-rich transmembrane protein 3 (Prrt3), mRNA	0.4042	0.0386
chr6	+	113604243	Irak2	Mus musculus interleukin-1 receptor-associated kinase 2 (Irak2), mRNA	0.3034	1.2742
chr6	-	115207967	Timp4	Mus musculus tissue inhibitor of metalloproteinase 4 (Timp4), mRNA	1.1752	0.2749
chr6	-	115694749	Tmem40	Mus musculus transmembrane protein 40 (Tmem40), mRNA	0.3151	0.223
chr6	-	117899177	Fxyd4	Mus musculus FXYD domain-containing ion transport regulator 4 (Fxyd4), mRNA	0.8679	0.1381
chr6	+	117976427	Rasgef1a	PREDICTED: Mus musculus RasGEF domain family, member 1A (Rasgef1a), mRNA	0.4602	0.1428
Cluster 5						

Chr	Strand	Start	Symbol	Gene name	Log ratio Oct4λVP2/ Oct4	Score
chr6	-	120896968	Mical3	Mus musculus microtubule associated monoxygenase, calponin and LIM domain containing 3 (Mical3), mRNA	0.4346	0.1144
chr6	+	121265845	Slc6a13	Mus musculus solute carrier family 6 (neurotransmitter transporter, GABA), member 13 (Slc6a13), mRNA	0.532	0.0937
chr6	+	121600994	A2m	Mus musculus alpha-2-macroglobulin (A2m), mRNA	1.2839	0.1933
chr6	-	122571018	Gdf3	Mus musculus growth differentiation factor 3 (Gdf3), mRNA	0.3699	0.1341
chr6	+	122592028	Dppa3	Mus musculus developmental pluripotency-associated 3 (Dppa3), mRNA	0.6708	0.3883
chr6	+	122673107	Nanog	PREDICTED: Mus musculus Nanog homeobox (Nanog), mRNA	0.4882	0.4754
chr6	+	122887466	Clec4a1	Mus musculus C-type lectin domain family 4, member a1 (Clec4a1), mRNA	0.6313	0.5354
chr6	+	122918133	Clec4a3	Mus musculus C-type lectin domain family 4, member a3 (Clec4a3), mRNA	0.7521	1.678
Cluster 6						
chr6	+	134947410	Apold1	PREDICTED: Mus musculus apolipoprotein L domain containing 1 (Apold1), mRNA	0.9952	0.1094
chr6	+	135031343	Gprc5a	Mus musculus G protein-coupled receptor, family C, group 5, member A (Gprc5a), mRNA	0.5753	0.1223
chr6	-	135103211	Hebp1	Mus musculus heme binding protein 1 (Hebp1), mRNA	0.4482	0.1034
chr6	-	135203021	Gsg1	Mus musculus germ cell-specific gene 1 (Gsg1), mRNA	0.8275	0.1894
chr6	-	135274821	Pbp2	Mus musculus phosphatidylethanolamine binding protein 2 (Pbp2), mRNA	0.8911	2.2829
chr6	+	135328237	Emp1	Mus musculus epithelial membrane protein 1 (Emp1), mRNA	0.949	2.4865
Cluster 7						
chr7	+	43219260	Siglecf	Mus musculus sialic acid binding Ig-like lectin F (Siglecf), mRNA	0.7657	0.1029
chr7	+	43295589	Lim2	Mus musculus lens intrinsic membrane protein 2 (Lim2), mRNA	0.8849	0.171
chr7	+	43311991	Etfb	Mus musculus electron transferring flavoprotein, beta polypeptide (Etfb), mRNA	0.4154	0.1713
chr7	+	43502748	Ceacam18	Mus musculus CEA-related cell adhesion molecule 1 (Ceacam18), mRNA	0.456	0.1266

Chr	Strand	Start	Symbol	Gene name	Log ratio Oct4λVP2/ Oct4	Score
chr7	+	43580608	Klk13	Mus musculus kallikrein 13 (Klk13), mRNA	0.3513	1.5903
chr7	+	43642658	Klk11	Mus musculus kallikrein 11 (Klk11), mRNA	0.7766	1.662
chr7	+	43649081	Klk10	Mus musculus kallikrein 10 (Klk10), mRNA	0.3525	1.3954
chr7	+	43834792	Klk1b1	Mus musculus kallikrein 1-related peptidase b1 (Klk1b1), mRNA	0.4043	1.4099
chr7	+	43880718	Klk1b26	Mus musculus kallikrein 1-related petidase b26 (Klk1b26), mRNA	0.5879	1.6478
chr7	+	44056304	Klk1b24	Mus musculus kallikrein 1-related peptidase b24 (Klk1b24), mRNA	0.3569	1.1982
chr7	-	44961034	Fcgrt	Mus musculus Fc receptor, IgG, alpha chain transporter (Fcgrt), mRNA	0.6001	0.3509
chr7	-	45101673	Cd37	Mus musculus CD37 antigen (Cd37), mRNA	0.6842	0.4157
chr7	-	45700821	Grin2d	Mus musculus glutamate receptor, ionotropic, NMDA2D (epsilon 4) (Grin2d), mRNA	0.4554	0.295
chr7	+	46244485	Myod1	Mus musculus myogenic differentiation 1 (Myod1), mRNA	0.3394	0.2227
chr7	+	46619874	Saa2	Mus musculus serum amyloid A 2 (Saa2), mRNA	0.5865	0.3214
chr7	-	46945816	Ptpn5	Mus musculus protein tyrosine phosphatase, non-receptor type 5 (Ptpn5), mRNA	0.4546	0.2733
Cluster 8						
chr8	-	125310857	Cyba	Mus musculus cytochrome b-245, alpha polypeptide (Cyba), mRNA	0.3059	0.1146
chr8	-	125340293	Snai3	Mus musculus snail homolog 3 (Drosophila) (Snai3), mRNA	0.4777	0.1133
chr8	-	125460724	Aprt	Mus musculus adenine phosphoribosyl transferase (Aprt), mRNA	0.3959	0.1733
chr8	-	125464327	Galns	Mus musculus galactosamine (N-acetyl)-6-sulfate sulfatase (Galns), mRNA	0.3551	0.1812
chr8	-	125515261	Cbfa2t3h	Mus musculus core-binding factor, runt domain, alpha subunit 2, translocated to, 3 homolog (human) (Cbfa2t3h), mRNA	0.4727	1.9641
chr8	-	126026018	Sult5a1	Mus musculus sulfotransferase family 5A, member 1 (Sult5a1), mRNA	0.4196	0.6186
chr8	+	126064817	Dpep1	Mus musculus dipeptidase 1 (renal) (Dpep1), mRNA	0.365	0.6649
chr8	+	126293171	Mc1r	Mus musculus melanocortin 1 receptor (Mc1r), mRNA	0.3715	0.4787

Chr	Strand	Start	Symbol	Gene name	Log ratio Oct4λVP2/ Oct4	Score
chr8	+	126297513	Tubb3	Mus musculus tubulin, beta 3 (Tubb3), mRNA	0.6688	0.5874
chr9	+	77894875	Ick	Mus musculus intestinal cell kinase (Ick), mRNA	0.3084	0.0771
chr9	+	78016339	Gsta1	Mus musculus glutathione S-transferase, alpha 1 (Ya) (Gsta1), mRNA	0.727	0.0882
chr9	+	78042812	Gsta1	Mus musculus glutathione S-transferase, alpha 1 (Ya) (Gsta1), mRNA	0.727	0.0921
chr9	+	78075611	Gsta2	Mus musculus glutathione S-transferase, alpha 2 (Yc2) (Gsta2), mRNA	0.4538	0.0992
chr9	+	78113713	Omt2b	Mus musculus oocyte maturation, beta (Omt2b), mRNA	0.3748	2.368
chr9	-	78116701	Gsta2	Mus musculus glutathione S-transferase, alpha 2 (Yc2) (Gsta2), mRNA	0.4538	5.4531
Cluster 9						
chr10	-	127912690	Mbc2	Mus musculus membrane bound C2 domain containing protein (Mbc2), mRNA	0.4804	0.0472
chr10	-	127970471	Erbb3	Mus musculus v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian) (Erbb3), mRNA	0.4644	0.0661
chr10	+	128193858	Mmp19	Mus musculus matrix metallopeptidase 19 (Mmp19), mRNA	0.4102	0.0674
chr10	+	128303937	Cd63	Mus musculus Cd63 antigen (Cd63), mRNA	0.476	0.2767
chr10	-	128361199	Mettl7b	Mus musculus methyltransferase like 7B (Mettl7b), mRNA	0.4484	1.0164
Cluster 10						
chr11	-	94053687	Wfikkn2	Mus musculus WAP, follistatin/kazal, immunoglobulin, kunitz and netrin domain containing 2 (Wfikkn2), mRNA	0.3786	0.0243
chr11	-	94159384	Abcc3	Mus musculus ATP-binding cassette, sub-family C (CFTR/MRP), member 3 (Abcc3), mRNA	0.5861	0.0394
chr11	-	94224460	Cacna1g	Mus musculus calcium channel, voltage-dependent, T type, alpha 1G subunit (Cacna1g), mRNA	0.6518	0.0677
chr11	-	94294993	Spata20	Mus musculus spermatogenesis associated 20 (Spata20), mRNA	0.3924	0.0734
chr11	+	95840634	Gip	Mus musculus gastric inhibitory polypeptide (Gip), mRNA	0.4817	0.2788

Chr	Strand	Start	Symbol	Gene name	Log ratio Oct4λVP2/ Oct4	Score
chr11	-	95915001	Ndp52	PREDICTED: Mus musculus nuclear domain 10 protein 52, transcript variant 4 (Ndp52), mRNA	0.4744	0.2946
chr11	+	95950308	Ttll6	Mus musculus tubulin tyrosine ligase-like family, member 6 (Ttll6), mRNA	0.4031	0.2785
chr11	+	96017396	LOC432593	PREDICTED: Mus musculus hypothetical gene supported by AK078606 (LOC432593), mRNA	0.534	0.2654
chr11	+	96115044	Hoxb6	Mus musculus homeo box B6 (Hoxb6), mRNA	0.5302	1.7663
chr11	+	97178543	Socs7	Mus musculus suppressor of cytokine signaling 7 (Socs7), mRNA	0.3578	0.364
chr11	-	97531246	Pip5k2b	Mus musculus phosphatidylinositol-4-phosphate 5-kinase, type II, beta (Pip5k2b), mRNA	0.3156	0.2708
chr11	+	98164479	Ppp1r1b	Mus musculus protein phosphatase 1, regulatory (inhibitor) subunit 1B (Ppp1r1b), mRNA	0.3208	0.1917
chr11	-	98584012	Nr1d1	Mus musculus nuclear receptor subfamily 1, group D, member 1 (Nr1d1), mRNA	0.3803	0.1543
chr11	-	98798269	Gjc1	Mus musculus gap junction membrane channel protein chi 1 (Gjc1), mRNA	0.4051	0.2197
chr11	-	98880560	Tns4	Mus musculus tensin 4 (Tns4), mRNA	0.5247	0.2885
Cluster 11						
chr13	+	3837001	Calm4	Mus musculus calmodulin 4 (Calm4), mRNA	0.881	-0.0049
chr13	+	3853513	Calm5	Mus musculus calmodulin 5 (Calm5), mRNA	1.3542	-0.0066
chr13	+	3923937	Tubal3	Mus musculus tubulin, alpha-like 3 (Tubal3), mRNA	0.3564	-0.0065
chr13	-	4131860	Akr1c18	Mus musculus aldo-keto reductase family 1, member C18 (Akr1c18), mRNA	0.6991	0.7956
chr13	+	4190429	Akr1c13	Mus musculus aldo-keto reductase family 1, member C13 (Akr1c13), mRNA	0.3627	2.0674
chr13	-	4267417	Akr1c12	Mus musculus aldo-keto reductase family 1, member C12 (Akr1c12), mRNA	0.3743	1.5205
Cluster 12						

Chr	Strand	Start	Symbol	Gene name	Log ratio Oct4λVP2/ Oct4	Score
chr13	-	32849524	Serpinb1a	Mus musculus serine (or cysteine) peptidase inhibitor, clade B, member 1a (Serpinb1a), mRNA	0.5608	0.0341
chr13	-	32888861	Serpinb1c	Mus musculus serine (or cysteine) peptidase inhibitor, clade B, member 1c (Serpinb1c), mRNA	0.4394	0.0735
chr13	+	32972673	Serpinb6b	Mus musculus serine (or cysteine) peptidase inhibitor, clade B, member 6b (Serpinb6b), mRNA	0.6722	0.122
chr13	+	33010714	Serpinb9	Mus musculus serine (or cysteine) peptidase inhibitor, clade B, member 9 (Serpinb9), mRNA	0.3426	0.2485
chr13	+	33034866	Serpinb9b	Mus musculus serine (or cysteine) peptidase inhibitor, clade B, member 9b (Serpinb9b), mRNA	0.5868	2.8076
chr13	+	33086037	Serpinb1b	Mus musculus serine (or cysteine) peptidase inhibitor, clade B, member 1b (Serpinb1b), mRNA	1.3932	3.4834
Cluster 13						
chr14	-	49687359	Ttc5	Mus musculus tetratricopeptide repeat domain 5 (Ttc5), mRNA	0.3061	0.0997
chr14	-	49828426	Osgep	Mus musculus O-sialoglycoprotein endopeptidase (Osgep), mRNA	0.3798	0.1017
chr14	+	49846908	Apex1	Mus musculus apurinic/apyrimidinic endonuclease 1 (Apex1), mRNA	0.435	0.1113
chr14	+	49866252	Pnp	Mus musculus purine-nucleoside phosphorylase (Pnp), mRNA	0.3587	0.1152
chr14	+	49877928	Pnp	Mus musculus purine-nucleoside phosphorylase (Pnp), mRNA	0.3587	1.9293
chr14	+	50013027	Ang1	Mus musculus angiogenin, ribonuclease A family, member 1 (Ang1), mRNA	0.5193	2.2226
chr14	-	50117436	Ang2	Mus musculus angiogenin, ribonuclease A family, member 2 (Ang2), mRNA	0.5315	1.6288
chr14	+	50283706	LOC638695	Mus musculus similar to Spetex-2C protein (LOC638695), mRNA	0.7666	1.2144
chr14	+	50474740	LOC434459	Mus musculus similar to RIKEN cDNA 4930503E14 (LOC434459), mRNA	0.7675	0.9864
chr14	-	50685828	Ang4	Mus musculus angiogenin, ribonuclease A family, member 4 (Ang4), mRNA	0.3178	0.8629
chr14	+	50721996	AY358078	Mus musculus cDNA sequence AY358078 (AY358078), mRNA	1.0044	1.1207
chr14	+	51032654	Rpgrip1	Mus musculus retinitis pigmentosa GTPase regulator interacting protein 1 (Rpgrip1), mRNA	0.3817	0.8647

Chr	Strand	Start	Symbol	Gene name	Log ratio Oct4λVP2/ Oct4	Score
chr14	-	51082364	Supt16h	Mus musculus suppressor of Ty 16 homolog (S. cerevisiae) (Supt16h), mRNA	0.3529	0.9296
chr14	-	51233128	Sall2	Mus musculus sal-like 2 (Drosophila) (Sall2), mRNA	0.3047	0.863
chr14	-	53193482	Dad1	Mus musculus defender against cell death 1 (Dad1), mRNA	0.4532	0.2021
chr14	+	53381113	Mrpl52	Mus musculus mitochondrial ribosomal protein L52 (Mrpl52), mRNA	0.3618	0.1579
chr14	+	53417248	Lrp10	Mus musculus low-density lipoprotein receptor-related protein 10 (Lrp10), mRNA	0.411	0.1613
chr14	-	53461382	Prmt5	Mus musculus protein arginine N-methyltransferase 5 (Prmt5), mRNA	0.4443	0.1773
chr14	-	53568317	Psmb5	Mus musculus proteasome (prosome, macropain) subunit, beta type 5 (Psmb5), mRNA	0.3664	1.0867
chr14	-	54015349	Zfhx2	Mus musculus zinc finger homeobox 2 (Zfhx2), mRNA	0.3556	0.6115
chr14	-	54052648	Ap1g2	Mus musculus adaptor protein complex AP-1, gamma 2 subunit (Ap1g2), mRNA	0.4312	0.6322
chr14	+	54176075	Dhrs2	Mus musculus dehydrogenase/reductase member 2 (Dhrs2), mRNA	0.3254	0.5381
chr14	+	54464555	Cpne6	Mus musculus copine VI (Cpne6), mRNA	0.3242	0.4023
chr14	+	54514102	Wdr23	Mus musculus WD repeat domain 23 (Wdr23), mRNA	0.3276	0.7074
chr14	+	55319138	Atp12a	Mus musculus ATPase, H+/K+ transporting, nongastric, alpha polypeptide (Atp12a), mRNA	0.6687	0.328
chr14	+	56752289	Sap18	Mus musculus Sin3-associated polypeptide 18 (Sap18), mRNA	0.4964	0.1663
chr14	-	56784169	Zdhhc20	Mus musculus zinc finger, DHHC domain containing 20 (Zdhhc20), mRNA	0.4585	0.1962
chr14	+	58155318	Rcbtb1	Mus musculus regulator of chromosome condensation (RCC1) and BTB (POZ) domain containing protein 1 (Rcbtb1), mRNA	0.798	0.151
chr14	-	58231021	Phf11	Mus musculus PHD finger protein 11 (Phf11), mRNA	2.2317	0.3196
chr14	-	58301312	D14Ertd668e	Mus musculus DNA segment, Chr 14, ERATO Doi 668, expressed (D14Ertd668e), mRNA	1.8165	0.749

Chr	Strand	Start	Symbol	Gene name	Log ratio Oct4λVP2/ Oct4	Score
chr14	-	58356118	Setdb2	PREDICTED: Mus musculus SET domain, bifurcated 2, transcript variant 2 (Setdb2), mRNA	1.8408	0.9091
chr14	+	58395072	Cab39l	Mus musculus calcium binding protein 39-like (Cab39l), mRNA	1.5962	6.9097
chr14	-	58513496	Cdadc1	PREDICTED: Mus musculus cytidine and dCMP deaminase domain containing 1, transcript variant 15 (Cdadc1), mRNA	1.3971	6.2889
chr14	+	58579390	Tmem46	Mus musculus transmembrane protein 46 (Tmem46), mRNA	1.5531	5.9003
chr14	-	58601631	Atp8a2	Mus musculus ATPase, aminophospholipid transporter-like, class I, type 8A, member 2 (Atp8a2), mRNA	0.4476	5.5678
Cluster 14						
chr15	-	74575097	Lynx1	Mus musculus Ly6/neurotoxin 1 (Lynx1), mRNA	0.508	0.0175
chr15	-	74624127	Ly6k	Mus musculus lymphocyte antigen 6 complex, locus K (Ly6k), mRNA	0.4581	0.0243
chr15	-	74822131	Ly6a	Mus musculus lymphocyte antigen 6 complex, locus A (Ly6a), mRNA	0.9588	0.0366
chr15	-	74872267	Ly6c	Mus musculus lymphocyte antigen 6 complex, locus C (Ly6c), mRNA	1.0214	0.0776
chr15	-	74901895	Ly6c	Mus musculus lymphocyte antigen 6 complex, locus C (Ly6c), mRNA	1.0214	2.4282
chr15	+	75531534	Rhpn1	Mus musculus rhophilin, Rho GTPase binding protein 1 (Rhpn1), mRNA	0.3252	0.8342
chr15	-	76439636	Slc39a4	Mus musculus solute carrier family 39 (zinc transporter), member 4 (Slc39a4), mRNA	0.3617	0.4561
Cluster 15						
chr17	+	33162394	Myo1f	Mus musculus myosin IF (Myo1f), mRNA	1.0917	0.0943
chr17	+	33984846	Btnl1	Mus musculus butyrophilin-like 1 (Btnl1), mRNA	0.9689	0.1227
chr17	+	34006550	BC051142	Mus musculus cDNA sequence BC051142 (BC051142), mRNA	0.6292	0.1391
chr17	+	34172264	Notch4	Mus musculus Notch gene homolog 4 (Drosophila) (Notch4), mRNA	0.6271	0.22
chr17	-	34633552	Ng23	Mus musculus Ng23 protein (Ng23), mRNA	0.7847	0.5576
chr17	+	34684986	Ly6g6e	Mus musculus lymphocyte antigen 6 complex, locus G6E (Ly6g6e), mRNA	0.5831	1.0264

Chr	Strand	Start	Symbol	Gene name	Log ratio Oct4λVP2/ Oct4	Score
chr17	-	34737052	Apom	Mus musculus apolipoprotein M (Apom), mRNA	0.643	0.8945

Table 3.5- List of the 147 genes upregulated in Oct4λVP2 cell line, which were identified as possible Oct4 targets.

The published lists are from studies by Matoba et al 2006, Loh et al 2006. Kim et al 2008, and Sharov et al 2008.
Pink colour represents genes tat are common to all the lists, Purple colour represents genes that are common to 3 lists, and the green colour represents genes that are common to two lists.

Symbol	Gene name					
Acp6	acid phosphatase 6, lysophosphatidic	target	target	target	target	target
Fbxo15	F-box protein 15	target	target	target	target	target
Nanog	Nanog homeobox		target	target	target	target
Ptch1	patched homolog 1		target	target	target	target
Tcfcp2l1	transcription factor CP2-like 1		target	target	target	target
Uhrf2	ubiquitin-like, containing PHD and RING finger domains 2		target	target	target	target
Ctnnal1	catenin (cadherin associated protein), alpha-like 1	target	target	target	target	
Nfatc2ip	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 interacting protein	target	target	target	target	
Tdh	L-threonine dehydrogenase	target	target	target	target	
Trh	thyrotropin releasing hormone	target	target	target	target	
Upp1	uridine phosphorylase 1	target	target	target	target	
Gadd45g	growth arrest and DNA-damage-inducible 45 gamma	target	target		target	target
Plscr1	phospholipid scramblase 1	target	target		target	target
Pax6	paired box gene 6			target	target	target
Sfrp1	secreted frizzled-related protein 1			target	target	target
Aff1	AF4/FMR2 family, member 1	target	target	target		
Ptges	prostaglandin E synthase	target	target	target		
Rpia	ribose 5-phosphate isomerase A	target	target	target		
Slc23a2	solute carrier family 23 (nucleobase transporters), member 2	target	target	target		

Symbol	Gene name					
Snai1	snail homolog 1 (Drosophila)	target	target	target		
Sap18	Sin3-associated polypeptide 18		target	target		
Smoc2	SPARC related modular calcium binding 2		target	target		
Tcfeb	transcription factor EB		target	target		
Cdyl	chromodomain protein, Y chromosome-like	target	target		target	
Manba	mannosidase, beta A, lysosomal	target	target		target	
Ms4a10	membrane-spanning 4-domains, subfamily A, member 10	target	target		target	
Pdcl2	phosducin-like 2	target	target		target	
Rage	renal tumor antigen	target	target		target	
Etv1	ets variant gene 1		target		target	
Ly6c1	lymphocyte antigen 6 complex, locus C1		target		target	
Prmt8	protein arginine N-methyltransferase 8	target	target			target
Lrrn2	leucine rich repeat protein 2, neuronal		target			target
Klf2	Kruppel-like factor 2 (lung)	target			target	target
Cd38	CD38 antigen				target	target
D14Ertd436e	DNA segment, Chr 14, ERATO Doi 436, expressed				target	target
Dppa3	developmental pluripotency-associated 3				target	target
Grsf1	G-rich RNA sequence binding factor 1				target	target
Itpka	inositol 1,4,5-trisphosphate 3-kinase A				target	target
Nucks1	nuclear casein kinase and cyclin-dependent kinase substrate 1				target	target
Ror1	receptor tyrosine kinase-like orphan receptor 1				target	target
Tmem8	transmembrane protein 8 (five membrane-spanning domains)				target	target
Vegfc	vascular endothelial growth factor C				target	target
Msc	musculin			target	target	
Myof	myoferlin			target	target	
Spp1	secreted phosphoprotein 1			target	target	

Symbol	Gene name	Loh 2006 Oct4-siRNA	Loh 2006 ChIP-Oct4	Matoba 2006	Sharov 2008	Kim 2008
Hmga2	high mobility group AT-hook 2	target	target			
Vcan	Versican core protein precursor (Large fibroblast proteoglycan) (Chondroitin sulfate proteoglycan core protein 2) (PG-M)	target	target			
Zbtb10	zinc finger and BTB domain containing 10	target		target		
Calcoco2	calcium binding and coiled-coil domain 2	target			target	
Id3	inhibitor of DNA binding3	target			target	
Mical1	microtubule associated monoxygenase, calponin and LIM domain containing 1 [	target			target	
Morc1	microrchidia 1	target			target	
Tcea3	transcription elongation factor A (SII), 3	target				target
Erbb3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian	target				
Ets2	E26 avian leukemia oncogene 2, 3' domain	target				
Irgq	immunity-related GTPase family, Q	target				
Prickle1	prickle like 1 (Drosophila)	target				
Abca4	ATP-binding cassette, sub-family A (ABC1), member 4		target			
Adam19	a disintegrin and metallopeptidase domain 19 (meltrin beta)		target			
Adcy5	adenylate cyclase 5		target			
Adrb3	adrenergic receptor, beta 3		target			
Atoh8	atonal homolog 8 (Drosophila)		target			
Atp8a2	ATPase, aminophospholipid transporter-like, class I, type 8A, member 2		target			
Cd63	melanoma 1 antigen		target			
Clnk	cytokine-dependent hematopoietic cell linker		target			
Crtac1	cartilage acidic protein 1		target			
Cxcr3	chemokine (C-X-C motif) receptor 3		target			

Symbol	Gene name	Loh 2006 Oct4-siRNA	Loh 2006 ChIP-Oct4	Matoba 2006	Sharov 2008	Kim 2008
Ephb4	Eph receptor B4		target			
Foxn2	forkhead box N2		target			
Glis3	GLIS family zinc finger 3		target			
Hoxb6	homeo box B6 [		target			
Lbxcor1	ladybird homeobox 1 homolog (Drosophila) corepressor 1		target			
Lmx1b	LIM homeobox transcription factor 1 beta		target			
Lrig1	leucine-rich repeats and immunoglobulin-like domains 1		target			
Ncam1	neural cell adhesion molecule 1		target			
Nfix	nuclear factor I/X		target			
Odz2	odd Oz/ten-m homolog 2 (Drosophila)		target			
Ppfia2	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 2		target			
Ptprj	protein tyrosine phosphatase, receptor type, J		target			
Rhbg	Rhesus blood group-associated B glycoprotein		target			
Sez6	seizure related gene 6		target			
Slc16a9	solute carrier family 16 (monocarboxylic acid transporters), member 9		target			
Tnfrsf8	tumor necrosis factor receptor superfamily, member 8		target			
Tox	thymocyte selection-associated high mobility group box		target			
Ttll6	tubulin tyrosine ligase-like family, member 6		target			
Unc5cl	unc-5 homolog C (C. elegans)-like		target			
Wdfy3	WD repeat and FYVE domain containing 3		target			
Yes1	Yamaguchi sarcoma viral (v-yes) oncogene homolog 1		target			
Dpysl3	dihydropyrimidinase-like 3			target		
Kitl	kit ligand			target		
Zc3hav1	zinc finger CCCH type, antiviral 1			target		

Symbol	Gene name	Loh 2006 Oct4-siRNA	Loh 2006 ChIP-Oct4	Matoba 2006	Sharov 2008	Kim 2006
Acoxl	acyl-Coenzyme A oxidase-like				target	
Ak7	adenylate kinase 7				target	
BC032203	BC032203 cDNA sequence BC032203				target	
Bud31	BUD31 homolog (yeast)				target	
Cdc42ep3	CDC42 effector protein (Rho GTPase binding) 3				target	
Cidea	cell death-inducing DNA fragmentation factor, alpha subunit-like effector A				target	
Dppa2	developmental pluripotency associated 2				target	
Emp1	epithelial membrane protein 1				target	
Fancm	Fanconi anemia, complementation group M				target	
Gjc1	gap junction protein, gamma 1				target	
Gjc1	gap junction protein, gamma 1				target	
Gnpnat1	glucosamine-phosphate N-acetyltransferase 1				target	
Gprc5b	G protein-coupled receptor, family C, group 5, member B [				target	
Htra1	HtrA serine peptidase 1				target	
Igfbp2	insulin-like growth factor binding protein 2				target	
Kit	kit oncogene				target	
Lgmn	legumain				target	
Mesdc1	mesoderm development candidate 1				target	
Mylpf	myosin light chain, phosphorylatable, fast skeletal muscle				target	
Ng23	Ng23 protein				target	
Pcolce2	procollagen C-endopeptidase enhancer 2				target	
Pla2g4e	phospholipase A2, group IVE				target	
Prtg	protogenin homolog (Gallus gallus)				target	
Rab27a	RAB27A, member RAS oncogene family				target	
Scap	SREBF chaperone				target	

Symbol	Gene name	Loh 2006 Oct4-siRNA	Loh 2006 ChIP-Oct4	Matoba 2006	Sharov 2008	Kim 2008
Shf	Src homology 2 domain containing F				target	
Slc15a1	solute carrier family 15 (oligopeptide transporter), member 1				target	
Slc39a4	solute carrier family 39 (zinc transporter), member 4				target	
Tfpi	tissue factor pathway inhibitor				target	
Tmem20	transmembrane protein 20				target	
Tspan17	tetraspanin 17				target	
Ulk1	Unc-51 like kinase 1 (C. elegans)				target	
Utrn	utrophin				target	
Zfp459	zinc finger protein 459				target	
Adamts4	a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 4					target
Ap3b2	adaptor-related protein complex 3, beta 2 subunit					target
Atoh1	atonal homolog 1 (Drosophila)					target
Cdk5r2	cyclin-dependent kinase 5, regulatory subunit 2 (p39)					target
Cdx1	caudal type homeo box 1					target
Cfc1	cripto, FRL-1, cryptic family 1					target
Cst3	cystatin C					target
Edn2	endothelin 2					target
Evx1	even skipped homeotic gene 1 homolog					target
Fam26e	family with sequence similarity 26, member E					target
Fn1	fibronectin 1					target
Gjb3	gap junction protein, beta 3					target
Ildr1	immunoglobulin-like domain containing receptor 1					target
Lefty2	left-right determination factor 2					target
Msln	: mesothelin					target
Nkx2-2	NK2 transcription factor related, locus 2					target

Symbol	Gene name					
Nrarp	Notch-regulated ankyrin repeat protein					target
Osr2	odd-skipped related 2 (Drosophila)					target
Pim1	proviral integration site 1					target
Ptger2	prostaglandin E receptor 2 (subtype EP2)					target
Rundc3a	RUN domain containing 3A					target
Shisa2	shisa homolog 2 (Xenopus laevis)					target
Slc12a7	solute carrier family 12, member 7					target
Sox13	SRY-box containing gene 13					target
Spic	Spi-C transcription factor (Spi-1/PU.1 related)					target
Stc2	stanniocalcin 2					target

Table 3.6- Gene ontology of the 147 genes upregulated in Oct4λVP2 cell line which were identified as possible Oct4 targets.

The genes identified are involved in cell fate commitment, growth, proliferation, differentiation, Kinase activities, and developmental process.

Annotation	p	FDR	Gene symbols
growth	0	0	Adrb3,Emp1,Htra1,Igfbp2,Lefty2,Lgmn,Ptch1,Tcfcp2l1,
negative regulation of growth	0	0	Adrb3,Lgmn,Ptch1,
growth factor activity	0.0012	0.006	Acoxl,Kitl,Lefty2,Spp1,Vegfc,
positive regulation of cell proliferation	0.0092	0.03	Kit,Kitl,Osr2,Spp1,Vegfc,
cell fate commitment	0.0026	0.011	Atoh1,Nanog,Nkx2-2,Pax6,
regulation of cell differentiation	0	0	Atoh1,Kitl,Nanog,Nkx2-2,Pax6,Spp1,Ulk1,
positive regulation of cell differentiation	0	0	Atoh1,Kitl,Nkx2-2,Pax6,
morphogenesis of an epithelium	0	0	Grsf1,Id3,Pax6,Ptch1,Tcfcp2l1,Vegfc,
neuron migration	0.0004	0.003	Atoh1,Lmx1b,Pax6,
axon guidance	0.0076	0.027	Atoh1,Etv1,Pax6,
transcription factor complex	0	0	Cdx1,Foxn2,Hoxb6,Lbxcor1,Lmx1b,Pax6,Sap18,Sox13,Spic, Tcfcp2l1,Tcfeb,
positive regulation of transcription	0.0011	0.006	Aff1,Atoh1,Etv1,Glis3,Klf2,Nkx2-2,Pax6,Tcfeb,
embryonic morphogenesis	0.0007	0.004	Atoh1,Grsf1,Lmx1b,Osr2,Ptch1,Vegfc,
dorsal/ventral pattern formation	0	0	Lmx1b,Pax6,Ptch1,
pattern specification process	0	0	Cdx1,Grsf1,Hoxb6,Lmx1b,Nanog,Pax6,Ptch1,Sfrp1,
regionalization	0.001	0.005	Grsf1,Lmx1b,Pax6,Ptch1,Sfrp1,
heart development	0.0041	0.016	Adam19,Gjc1,Id3,Ptprj,Vcan,

Annotation	p	FDR	Gene symbols
gland development	0.0065	0.024	Pax6,Ptch1,Tcfcp2l1,
chemotaxis	0.0013	0.006	Cxcr3,Kit,Spp1,Utrn,
locomotory behavior	0.0014	0.007	Adcy5,Cxcr3,Kit,Spp1,Trh,Utrn,
wound healing	0.0076	0.027	Fn1,Rab27a,Tfpi,
glycosaminoglycan binding	0	0	Fn1,Ncam1,Pcolce2,Ptch1,Vcan,
adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	0.0076	0.028	Gadd45g,Rab27a,Spp1,
enzyme linked receptor protein signaling pathway	0	0	Cidea,Clnk,Ephb4,Fn1,Htra1,Kit,Ptprj,Ulk1,Vegfc,
positive regulation of protein kinase activity	0.0127	0.041	Gadd45g,Kit,Kitl,
protein-tyrosine kinase activity	0.0015	0.007	Ephb4,Kit,Ncam1,Ror1,Yes1,
transmembrane receptor protein tyrosine kinase signaling pathway	0.0002	0.001	Clnk,Ephb4,Fn1,Kit,Ulk1,Vegfc,
transmembrane receptor protein kinase activity	0	0	Ephb4,Kit,Ncam1,Ror1,
symporter activity	0.0002	0.001	Slc12a7,Slc15a1,Slc16a9,Slc23a1,Slc23a2,
late endosome	0	0	Cd63,Gnpnat1,Lgmn,

Table 3.7- List of genes upregulated in Oct4λVP2 that are targets of either Oct4 only, Nanog only, or both.

The table includes regulation of the genes after Nanog or Oct4 siRNA treatment: UP means genes are upregulated after the siRNA treatment, Down means genes are downregulated after the siRNA treatment.
The columns labelled Nanog targets and Oct4 targets refer to Nanog and Oct4 targets identified in previous studies. Targets identified in each study were allocated a specific colour: Blue colour refers to targets from Loh et al 2006, purple colour refers to targets from Matoba et al 2006, Green colour refers to Kim et al 2008, and pink colour refers to sharov et al 2008. Targets identified in multiple studies are coloured in more than one colour (each colour representing the appropriate study the target was identified in).

Symbol	Gene name	Oct4 siRNA Nanog	Nanog targets	Oct4 targets	Oct4 siRNA Oct4
Calcoco2	calcium binding and coiled-coil domain 2	UP	target		UP
Prickle1	prickle like 1 (Drosophila)	UP	target		UP
Id3	inhibitor of DNA binding 3	UP			UP
Ms4a10	membrane-spanning 4-domains, subfamily A, member 10	UP		target	UP
Snai1	snail homolog 1 (Drosophila)	UP		target	UP
Ctnnal1	catenin (cadherin associated protein), alpha-like 1	UP	target	target	Down
Bud31	BUD31 homolog (yeast)	UP	target		
Tmem20	transmembrane protein 20	UP	target		
Adrb3	adrenergic receptor, beta 3	UP	target	target	
Cd38	CD38 antigen	UP	target	target	
D14Ertd436e	DNA segment, Chr 14, ERATO Doi 436, expressed	UP	target	target	
Dppa3	developmental pluripotency-associated 3	UP	target	target	
Gjb3	gap junction protein, beta 3	UP	target	target	
Aspa	aspartoacylase	UP	target		
Glrx	glutaredoxin	UP	target		

Symbol	Gene name	Loh2006 siRNA Nanog	Nanog chip	mOct4 chip	[illegible]
Gsta3	glutathione S-transferase, alpha 3	UP	target		
Klhl13	kelch-like 13 (Drosophila)	UP	target		
Slc25a20	solute carrier family 25 (mitochondrial carnitine/acylcarnitine translocase), member 20	UP	target		
Slco2a1	solute carrier organic anion transporter family, member 2a1	UP	target		
Mylpf	myosin light chain, phosphorylatable, fast skeletal muscle	UP			
Prtg	protogenin homolog (Gallus gallus)	UP			
Scap	SREBF chaperone	UP			
Grsf1	G-rich RNA sequence binding factor 1	UP		target	
Cryab	crystallin, alpha B	UP			
Echdc2	enoyl Coenzyme A hydratase domain containing 2	UP			
Eng	endoglin	UP			
Folr1	folate receptor 1 (adult)	UP			
Hist2h3c2	histone cluster 2, H3c2	UP			
Lxn	latexin	UP			
Mmrn2	multimerin 2	UP			
Pcdh8	protocadherin 8	UP			
Pramel6	preferentially expressed antigen in melanoma like 6	UP			
Prep	prolyl endopeptidase	UP			
Stmn3	stathmin-like 3	UP			
Tesk2	testis-specific kinase 2	UP			
Ttl	tubulin tyrosine ligase	UP			
Mical1	microtubule associated monoxygenase, calponin and LIM domain containing 1	Down	target		UP
Manba	mannosidase, beta A, lysosomal	Down		target	UP
Pdcl2	phosducin-like 2	Down		target	UP
Upp1	uridine phosphorylase 1	Down		target	UP

Symbol	Gene name	Lon2 siRNA Nanog	Nanog chip	mOct4 chip	[illegible]
Gjc1	gap junction protein, gamma 1	Down	target		
Crtac1	cartilage acidic protein 1	Down	target	target	
Fads2	fatty acid desaturase 2	Down	target		
Sin3a	transcriptional regulator, SIN3A (yeast)	Down	target		
Tubb2b	tubulin, beta 2B	Down	target		
BC032203	cDNA sequence BC032203	Down			
Igfbp2	insulin-like growth factor binding protein 2	Down			
Spp1	secreted phosphoprotein 1	Down			
Nanog		Down		target	
Ephb4	Eph receptor B4	Down		target	
Itpka	inositol 1,4,5-trisphosphate 3-kinase A	Down		target	
Cab39l	calcium binding protein 39-like	Down			
Car13	carbonic anhydrase 13	Down			
Idh2	isocitrate dehydrogenase 2 (NADP+), mitochondrial	Down			
Mov10	Moloney leukemia virus 10	Down			
Myo1f		Down			
Osgep	O-sialoglycoprotein endopeptidase	Down			
Pgm1	phosphoglucomutase 1	Down			
Pla2g10	phospholipase A2, group X	Down			
Scd1	stearoyl-Coenzyme A desaturase 1	Down			
Tmem40	transmembrane protein 40	Down			
Gadd45g	growth arrest and DNA-damage-inducible 45 gamma		target	target	UP
Zbtb10	zinc finger and BTB domain containing 10				UP
Hmga2	high mobility group AT-hook 2			target	UP
Erbb3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)				UP
Ets2	E26 avian leukemia oncogene 2, 3' domain		target		UP
Cdyl	chromodomain protein, Y chromosome-like		target	target	Down

Symbol	Gene name	[illegible]	[illegible]	[illegible]	[illegible]
Rage	renal tumor antigen		target	target	Down
Aff1	AF4/FMR2 family, member 1		target	target	Down
Ptges	prostaglandin E synthase		target	target	Down
Acp6	acid phosphatase 6, lysophosphatidic		target	target	Down
Fbxo15	F-box protein 15		target	target	Down
Plscr1	phospholipid scramblase 1		target	target	Down
Nfatc2ip	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 interacting protein			target	Down
Tdh	L-threonine dehydrogenase			target	Down
Trh	thyrotropin releasing hormone			target	Down
Prmt8	protein arginine N-methyltransferase 8			target	Down
Vcan	versican			target	Down
Klf2	Kruppel-like factor 2 (lung)			target	Down
Tcea3	transcription elongation factor A (SII), 3			target	Down
Morc1	microrchidia 1			target	Down
Rpia	ribose 5-phosphate isomerase A			target	Down
Slc23a2	solute carrier family 23 (nucleobase transporters), member 2			target	Down
Irgq	immunity-related GTPase family, Q				Down

REFERENCES

References

Ambrosetti, D.C., Scholer, H.R., Dailey, L., and Basilico, C. (2000). Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. J Biol Chem *275*, 23387-23397.

Aoto, T., Saitoh, N., Ichimura, T., Niwa, H., and Nakao, M. (2006). Nuclear and chromatin reorganization in the MHC-Oct3/4 locus at developmental phases of embryonic stem cell differentiation. Dev Biol *298*, 354-367.

Arnold, S.J., and Robertson, E.J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nat Rev Mol Cell Biol *10*, 91-103.

Auernhammer, C.J., Bousquet, C., Chesnokova, V., and Melmed, S. (2000). SOCS proteins: modulators of neuroimmunoendocrine functions. Impact on corticotroph LIF signaling. Ann N Y Acad Sci *917*, 658-664.

Aurora, R., and Herr, W. (1992). Segments of the POU domain influence one another's DNA-binding specificity. Mol Cell Biol *12*, 455-467.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev *17*, 126-140.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., *et al.* (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol *8*, 532-538.

Babaie, Y., Herwig, R., Greber, B., Brink, T.C., Wruck, W., Groth, D., Lehrach, H., Burdon, T., and Adjaye, J. (2007). Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. Stem Cells *25*, 500-510.

Bardwell, L., and Shah, K. (2006). Analysis of mitogen-activated protein kinase activation and interactions with regulators and substrates. Methods *40*, 213-223.

Barnea, E., and Bergman, Y. (2000). Synergy of SF1 and RAR in activation of Oct-3/4 promoter. J Biol Chem *275*, 6608-6619.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Batlle-Morera, L., Smith, A., and Nichols, J. (2008). Parameters influencing derivation of embryonic stem cells from murine embryos. Genesis *46*, 758-767.

Beddington, R.S., and Robertson, E.J. (1999). Axis development and early asymmetry in mammals. Cell *96*, 195-209.

Belting, H.G., Hauptmann, G., Meyer, D., Abdelilah-Seyfried, S., Chitnis, A., Eschbach, C., Soll, I., Thisse, C., Thisse, B., Artinger, K.B., *et al.* (2001). spiel ohne grenzen/pou2 is required during establishment of the zebrafish midbrain-hindbrain boundary organizer. Development *128*, 4165-4176.

Ben-Shushan, E., Pikarsky, E., Klar, A., and Bergman, Y. (1993). Extinction of Oct-3/4 gene expression in embryonal carcinoma x fibroblast somatic cell hybrids is accompanied by changes in the methylation status, chromatin structure, and transcriptional activity of the Oct-3/4 upstream region. Mol Cell Biol *13*, 891-901.

Ben-Shushan, E., Sharir, H., Pikarsky, E., and Bergman, Y. (1995). A dynamic balance between ARP-1/COUP-TFII, EAR-3/COUP-TFI, and retinoic acid receptor:retinoid X receptor heterodimers regulates Oct-3/4 expression in embryonal carcinoma cells. Mol Cell Biol *15*, 1034-1048.

Ben-Shushan, E., Thompson, J.R., Gudas, L.J., and Bergman, Y. (1998). Rex-1, a gene encoding a transcription factor expressed in the early embryo, is regulated via Oct-3/4 and Oct-6 binding to an octamer site and a novel protein, Rox-1, binding to an adjacent site. Mol Cell Biol *18*, 1866-1878.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. Cell *128*, 669-681.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bi, L., Okabe, I., Bernard, D.J., and Nussbaum, R.L. (2002). Early embryonic lethality in mice deficient in the p110beta catalytic subunit of PI 3-kinase. Mamm Genome *13*, 169-172.

Bi, L., Okabe, I., Bernard, D.J., Wynshaw-Boris, A., and Nussbaum, R.L. (1999). Proliferative defect and embryonic lethality in mice homozygous for a deletion in the p110alpha subunit of phosphoinositide 3-kinase. J Biol Chem *274*, 10963-10968.

Blom, N., Gammeltoft, S., and Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol *294*, 1351-1362.

Boiani, M., and Scholer, H.R. (2005). Regulatory networks in embryo-derived pluripotent stem cells. Nat Rev Mol Cell Biol *6*, 872-884.

Botfield, M.C., Jancso, A., and Weiss, M.A. (1992). Biochemical characterization of the Oct-2 POU domain with implications for bipartite DNA recognition. Biochemistry *31*, 5841-5848.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Boyer, L.A., Mathur, D., and Jaenisch, R. (2006a). Molecular control of pluripotency. Curr Opin Genet Dev *16*, 455-462.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., *et al.* (2006b). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Brachmann, S.M., Yballe, C.M., Innocenti, M., Deane, J.A., Fruman, D.A., Thomas, S.M., and Cantley, L.C. (2005). Role of phosphoinositide 3-kinase regulatory isoforms in development and actin rearrangement. Mol Cell Biol *25*, 2593-2606.

Brehm, A., Ohbo, K., Zwerschke, W., Botquin, V., Jansen-Durr, P., and Scholer, H.R. (1999). Synergism with germ line transcription factor Oct-4: viral oncoproteins share the ability to mimic a stem cell-specific activity. Mol Cell Biol *19*, 2635-2643.

Brehm, A., Ovitt, C.E., and Scholer, H.R. (1998). Oct-4: more than just a POUerful marker of the mammalian germline? Apmis *106*, 114-124; discussion 124-116.

Brickman, J.M., Clements, M., Tyrell, R., McNay, D., Woods, K., Warner, J., Stewart, A., Beddington, R.S., and Dattani, M. (2001). Molecular effects of novel mutations in Hesx1/HESX1 associated with human pituitary disorders. Development *128*, 5189-5199.

Brickman, J.M., Jones, C.M., Clements, M., Smith, J.C., and Beddington, R.S. (2000). Hex is a transcriptional repressor that contributes to anterior identity and suppresses Spemann organiser function. Development *127*, 2303-2315.

Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., *et al.* (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191-195.

Brook, F.A., and Gardner, R.L. (1997). The origin and efficient derivation of embryonic stem cells in the mouse. Proc Natl Acad Sci U S A *94*, 5709-5712.

Buehr, M., Meek, S., Blair, K., Yang, J., Ure, J., Silva, J., McLay, R., Hall, J., Ying, Q.L., and Smith, A. (2008). Capture of authentic embryonic stem cells from rat blastocysts. Cell *135*, 1287-1298.

Burdon, T., Stracey, C., Chambers, I., Nichols, J., and Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. Dev Biol *210*, 30-43.

Cadigan, K.M., and Nusse, R. (1997). Wnt signaling: a common theme in animal development. Genes Dev *11*, 3286-3305.

Campbell, M.E., Palfreyman, J.W., and Preston, C.M. (1984). Identification of herpes simplex virus DNA sequences which encode a trans-acting polypeptide responsible for stimulation of immediate early transcription. J Mol Biol *180*, 1-19.

Campbell, P.A., Perez-Iratxeta, C., Andrade-Navarro, M.A., and Rudnicki, M.A. (2007). Oct4 targets regulatory nodes to modulate stem cell function. PLoS One *2*, e553.

Cao, Y., Knochel, S., Donow, C., Miethe, J., Kaufmann, E., and Knochel, W. (2004). The POU factor Oct-25 regulates the Xvent-2B gene and counteracts terminal differentiation in Xenopus embryos. J Biol Chem *279*, 43735-43743.

Cao, Y., Siegel, D., and Knochel, W. (2006). Xenopus POU factors of subclass V inhibit activin/nodal signaling during gastrulation. Mech Dev *123*, 614-625.

Carey, M., Leatherwood, J., and Ptashne, M. (1990). A potent GAL4 derivative activates transcription at a distance in vitro. Science *247*, 710-712.

Carter, D.E., Robinson, J.F., Allister, E.M., Huff, M.W., and Hegele, R.A. (2005). Quality assessment of microarray experiments. Clin Biochem *38*, 639-642.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development *132*, 885-896.

Chamberlain, S.J., Yee, D., and Magnuson, T. (2008). Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. Stem Cells *26*, 1496-1505.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.

Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Dev Cell *10*, 615-624.

Chen, L., Yabuuchi, A., Eminli, S., Takeuchi, A., Lu, C.W., Hochedlinger, K., and Daley, G.Q. (2009). Cross-regulation of the Nanog and Cdx2 promoters. Cell Res *19*, 1052-1061.

Chen, X., Vega, V.B., and Ng, H.H. (2008a). Transcriptional regulatory networks in embryonic stem cells. Cold Spring Harb Symp Quant Biol *73*, 203-209.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., *et al.* (2008b). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.

Cheng, A.M., Saxton, T.M., Sakai, R., Kulkarni, S., Mbamalu, G., Vogel, W., Tortorice, C.G., Cardiff, R.D., Cross, J.C., Muller, W.J., *et al.* (1998). Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. Cell *95*, 793-803.

Chew, J.L., Loh, Y.H., Zhang, W., Chen, X., Tam, W.L., Yeap, L.S., Li, P., Ang, Y.S., Lim, B., Robson, P., *et al.* (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Mol Cell Biol *25*, 6031-6046.

Chou, Y.F., Chen, H.H., Eijpe, M., Yabuuchi, A., Chenoweth, J.G., Tesar, P., Lu, J., McKay, R.D., and Geijsen, N. (2008). The growth factor environment defines distinct pluripotent ground states in novel blastocyst-derived stem cells. Cell *135*, 449-461.

Clerc, R.G., Corcoran, L.M., LeBowitz, J.H., Baltimore, D., and Sharp, P.A. (1988). The B-cell-specific Oct-2 protein contains POU box- and homeo box-type domains. Genes Dev *2*, 1570-1581.

Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes Dev *22*, 746-755.

Constance Lane, M., Davidson, L., and Sheets, M.D. (2004). BMP antagonism by Spemann's organizer regulates rostral-caudal fate of mesoderm. Dev Biol *275*, 356-374.

Croston, G.E., Laybourn, P.J., Paranjape, S.M., and Kadonaga, J.T. (1992). Mechanism of transcriptional antirepression by GAL4-VP16. Genes Dev *6*, 2270-2281.

Cui, L., Johkura, K., Yue, F., Ogiwara, N., Okouchi, Y., Asanuma, K., and Sasaki, K. (2004). Spatial distribution and initial changes of SSEA-1 and other cell adhesion-related molecules on mouse embryonic stem cells before and during differentiation. J Histochem Cytochem *52*, 1447-1457.

Daheron, L., Opitz, S.L., Zaehres, H., Lensch, M.W., Andrews, P.W., Itskovitz-Eldor, J., and Daley, G.Q. (2004). LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. Stem Cells *22*, 770-778.

Ding, L., Paszkowski-Rogacz, M., Nitzsche, A., Slabicki, M.M., Heninger, A.K., de Vries, I., Kittler, R., Junqueira, M., Shevchenko, A., Schulz, H., *et al.* (2009). A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. Cell Stem Cell *4*, 403-415.

Do, J.T., and Scholer, H.R. (2009). Regulatory circuits underlying pluripotency and reprogramming. Trends Pharmacol Sci *30*, 296-302.

Emami, K.H., and Carey, M. (1992). A synergistic increase in potency of a multimerized VP16 transcriptional activation domain. Embo J *11*, 5005-5012.

Endoh, M., Endo, T.A., Endoh, T., Fujimura, Y., Ohara, O., Toyoda, T., Otte, A.P., Okano, M., Brockdorff, N., Vidal, M., *et al.* (2008). Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development *135*, 1513-1524.

Episkopou, V. (2005). SOX2 functions in adult neural stem cells. Trends Neurosci *28*, 219-221.

Ernst, M., and Jenkins, B.J. (2004). Acquiring signalling specificity from the cytokine receptor gp130. Trends Genet *20*, 23-32.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

Ezashi, T., Ghosh, D., and Roberts, R.M. (2001). Repression of Ets-2-induced transactivation of the tau interferon promoter by Oct-4. Mol Cell Biol *21*, 7883-7891.

Fauquier, T., Rizzoti, K., Dattani, M., Lovell-Badge, R., and Robinson, I.C. (2008). SOX2-expressing progenitor cells generate all of the major cell types in the adult mouse pituitary gland. Proc Natl Acad Sci U S A *105*, 2907-2912.

Faust, C., Lawson, K.A., Schork, N.J., Thiel, B., and Magnuson, T. (1998). The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. Development *125*, 4495-4506.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat Cell Biol *8*, 188-194.

Finney, M., Ruvkun, G., and Horvitz, H.R. (1988). The C. elegans cell lineage and differentiation gene unc-86 encodes a protein with a homeodomain and extended similarity to transcription factors. Cell *55*, 757-769.

Foygel, K., Choi, B., Jun, S., Leong, D.E., Lee, A., Wong, C.C., Zuo, E., Eckart, M., Reijo Pera, R.A., Wong, W.H., *et al.* (2008). A novel and critical role for Oct4 as a regulator of the maternal-embryonic transition. PLoS One *3*, e4109.

Fuhrmann, G., Chung, A.C., Jackson, K.J., Hummelke, G., Baniahmad, A., Sutter, J., Sylvester, I., Scholer, H.R., and Cooney, A.J. (2001). Mouse germline restriction of Oct4 expression by germ cell nuclear factor. Dev Cell *1*, 377-387.

Fujikura, J., Yamato, E., Yonemura, S., Hosoda, K., Masui, S., Nakao, K., Miyazaki Ji, J., and Niwa, H. (2002). Differentiation of embryonic stem cells is induced by GATA factors. Genes Dev *16*, 784-789.

Gammill, L.S., and Sive, H. (2001). otx2 expression in the ectoderm activates anterior neural determination and is required for Xenopus cement gland formation. Dev Biol *240*, 223-236.

Grigoryan, T., Wend, P., Klaus, A., and Birchmeier, W. (2008). Deciphering the function of canonical Wnt signals in development and disease: conditional loss- and gain-of-function mutations of beta-catenin in mice. Genes Dev *22*, 2308-2341.

Gu, P., Goodwin, B., Chung, A.C., Xu, X., Wheeler, D.A., Price, R.R., Galardi, C., Peng, L., Latour, A.M., Koller, B.H., *et al.* (2005). Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. Mol Cell Biol *25*, 3492-3505.

Guo, Y., Costa, R., Ramsey, H., Starnes, T., Vance, G., Robertson, K., Kelley, M., Reinbold, R., Scholer, H., and Hromas, R. (2002). The embryonic stem cell transcription factors Oct-4 and FoxD3 interact to regulate endodermal-specific promoter expression. Proc Natl Acad Sci U S A *99*, 3663-3667.

Han, K., and Manley, J.L. (1993a). Functional domains of the Drosophila Engrailed protein. Embo J *12*, 2723-2733.

Han, K., and Manley, J.L. (1993b). Transcriptional repression by the Drosophila even-skipped protein: definition of a minimal repression domain. Genes Dev *7*, 491-503.

Han, Y.M., Dahlberg, J., Lund, E., Manley, J.L., and Prives, C. (1991). SV40 T-antigen-binding sites within the 5'-flanking regions of human U1 and U2 genes. Gene *109*, 219-231.

Harry J. Kraft, S.M., Hessel A. Smits, Peter Hohenstein, Ester Piek, Qi Chen, Karen Artzt, and Everardus J. J. van Zoelen (1996). Oct-4 Regulates Alternative Platelet-derived Growth Factor α Receptor Gene Promoter in Human Embryonal Carcinoma Cells. J Biol Chem *271*, 12873-12878.

Hart, A.H., Hartley, L., Ibrahim, M., and Robb, L. (2004). Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. Dev Dyn *230*, 187-198.

Hauptmann, G., Belting, H.G., Wolke, U., Lunde, K., Soll, I., Abdelilah-Seyfried, S., Prince, V., and Driever, W. (2002). spiel ohne grenzen/pou2 is required for zebrafish hindbrain segmentation. Development *129*, 1645-1655.

Hay, D.C., Sutherland, L., Clark, J., and Burdon, T. (2004). Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. Stem Cells *22*, 225-235.

Herr, W., Sturm, R.A., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M., Ruvkun, G., *et al.* (1988). The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products. Genes Dev *2*, 1513-1516.

Hill, C.S. (2009). Nucleocytoplasmic shuttling of Smad proteins. Cell Res *19*, 36-46.

Hinkley, C.S., Martin, J.F., Leibham, D., and Perry, M. (1992). Sequential expression of multiple POU proteins during amphibian early development. Mol Cell Biol *12*, 638-649.

Howley, C., and Ho, R.K. (2000). mRNA localization patterns in zebrafish oocytes. Mech Dev *92*, 305-309.

Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev *23*, 837-848.

Huelsken, J., and Birchmeier, W. (2001). New aspects of Wnt signaling pathways in higher vertebrates. Curr Opin Genet Dev *11*, 547-553.

Humphrey, R.K., Beattie, G.M., Lopez, A.D., Bucay, N., King, C.C., Firpo, M.T., Rose-John, S., and Hayek, A. (2004). Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. Stem Cells *22*, 522-530.

Imagawa, M., Miyamoto, A., Shirakawa, M., Hamada, H., and Muramatsu, M. (1991). Stringent integrity requirements for both trans-activation and DNA-binding in a trans-activator, Oct3. Nucleic Acids Res *19*, 4503-4508.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. (2006). Dissecting self-renewal in stem cells with RNA interference. Nature *442*, 533-538.

Jacobson, E.M., Li, P., Leon-del-Rio, A., Rosenfeld, M.G., and Aggarwal, A.K. (1997). Structure of Pit-1 POU domain bound to DNA as a dimer: unexpected arrangement and flexibility. Genes Dev *11*, 198-212.

Jaynes, J.B., and O'Farrell, P.H. (1991). Active repression of transcription by the engrailed homeodomain protein. Embo J *10*, 1427-1433.

Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol *10*, 353-360.

Jirmanova, L., Afanassieff, M., Gobert-Gosse, S., Markossian, S., and Savatier, P. (2002). Differential contributions of ERK and PI3-kinase to the regulation of cyclin D1 expression and to the control of the G1/S transition in mouse embryonic stem cells. Oncogene *21*, 5515-5528.

Johnson, R., Teh, C.H., Kunarso, G., Wong, K.Y., Srinivasan, G., Cooper, M.L., Volta, M., Chan, S.S., Lipovich, L., Pollard, S.M., *et al.* (2008). REST regulates distinct transcriptional networks in embryonic and neural stem cells. PLoS Biol *6*, e256.

Kidder, B.L., Oseth, L., Miller, S., Hirsch, B., Verfaillie, C., and Coucouvanis, E. (2008). Embryonic stem cells contribute to mouse chimeras in the absence of detectable cell fusion. Cloning Stem Cells *10*, 231-248.

Kielman, M.F., Rindapaa, M., Gaspar, C., van Poppel, N., Breukel, C., van Leeuwen, S., Taketo, M.M., Roberts, S., Smits, R., and Fodde, R. (2002). Apc modulates embryonic stem-cell differentiation by controlling the dosage of beta-catenin signaling. Nat Genet *32*, 594-605.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049-1061.

Kinoshita, K., Ura, H., Akagi, T., Usuda, M., Koide, H., and Yokota, T. (2007). GABPalpha regulates Oct-3/4 expression in mouse embryonic stem cells. Biochem Biophys Res Commun *353*, 686-691.

Kirchhof, N., Carnwath, J.W., Lemme, E., Anastassiadis, K., Scholer, H., and Niemann, H. (2000). Expression pattern of Oct-4 in preimplantation embryos of different species. Biol Reprod *63*, 1698-1705.

Klemm, J.D., and Pabo, C.O. (1996). Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage. Genes Dev *10*, 27-36.

Klemm, J.D., Rould, M.A., Aurora, R., Herr, W., and Pabo, C.O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. Cell *77*, 21-32.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., *et al.* (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet *4*, e1000242.

Kunath, T., Arnaud, D., Uy, G.D., Okamoto, I., Chureau, C., Yamanaka, Y., Heard, E., Gardner, R.L., Avner, P., and Rossant, J. (2005). Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. Development *132*, 1649-1661.

Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development *134*, 2895-2902.

Labosky, P.A., Barlow, D.P., and Hogan, B.L. (1994). Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. Development *120*, 3197-3204.

Lavial, F., Acloque, H., Bertocchini, F., Macleod, D.J., Boast, S., Bachelard, E., Montillet, G., Thenot, S., Sang, H.M., Stern, C.D., *et al.* (2007). The Oct4 homologue PouV and Nanog regulate pluripotency in chicken embryonic stem cells. Development *134*, 3549-3563.

Lee, E.R., Murdoch, F.E., and Fritsch, M.K. (2007a). High histone acetylation and decreased polycomb repressive complex 2 member levels regulate gene specific transcriptional changes during early embryonic stem cell differentiation induced by retinoic acid. Stem Cells *25*, 2191-2199.

Lee, K.Y., Jeong, J.W., Wang, J., Ma, L., Martin, J.F., Tsai, S.Y., Lydon, J.P., and DeMayo, F.J. (2007b). Bmp2 is critical for the murine uterine decidual response. Mol Cell Biol *27*, 5468-5478.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., *et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Lenardo, M.J., Staudt, L., Robbins, P., Kuang, A., Mulligan, R.C., and Baltimore, D. (1989). Repression of the IgH enhancer in teratocarcinoma cells associated with a novel octamer factor. Science *243*, 544-546.

Lengner, C.J., Welstead, G.G., and Jaenisch, R. (2008). The pluripotency regulator Oct4: a role in somatic stem cells? Cell Cycle *7*, 725-728.

Levasseur, D.N., Wang, J., Dorschner, M.O., Stamatoyannopoulos, J.A., and Orkin, S.H. (2008). Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. Genes Dev *22*, 575-580.

Li, J.Y., Pu, M.T., Hirasawa, R., Li, B.Z., Huang, Y.N., Zeng, R., Jing, N.H., Chen, T., Li, E., Sasaki, H., *et al.* (2007). Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. Mol Cell Biol *27*, 8748-8759.

Li, P., Tong, C., Mehrian-Shai, R., Jia, L., Wu, N., Yan, Y., Maxson, R.E., Schulze, E.N., Song, H., Hsieh, C.L., *et al.* (2008). Germline competent embryonic stem cells derived from rat blastocysts. Cell *135*, 1299-1310.

Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S.Y., Qin, J., Wong, J., Cooney, A.J., Liu, D., *et al.* (2008). Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat Cell Biol *10*, 731-739.

Lim, C.Y., Tam, W.L., Zhang, J., Ang, H.S., Jia, H., Lipovich, L., Ng, H.H., Wei, C.L., Sung, W.K., Robson, P., *et al.* (2008). Sall4 regulates distinct transcription circuitries in different blastocyst-derived stem cell lineages. Cell Stem Cell *3*, 543-554.

Liu, L., Leaman, D., Villalta, M., and Roberts, R.M. (1997). Silencing of the gene for the alpha-subunit of human chorionic gonadotropin by the embryonic transcription factor Oct-3/4. Mol Endocrinol *11*, 1651-1658.

Liu, L., and Roberts, R.M. (1996). Silencing of the gene for the beta subunit of human chorionic gonadotropin by the embryonic transcription factor Oct-3/4. J Biol Chem *271*, 16683-16689.

Logan, C.Y., and Nusse, R. (2004). The Wnt signaling pathway in development and disease. Annu Rev Cell Dev Biol *20*, 781-810.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., *et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431-440.

Loh, Y.H., Zhang, W., Chen, X., George, J., and Ng, H.H. (2007). Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. Genes Dev *21*, 2545-2557.

Lu, D.P., Chandrakanthan, V., Cahana, A., Ishii, S., and O'Neill, C. (2004). Trophic signals acting via phosphatidylinositol-3 kinase are required for normal pre-implantation mouse embryo development. J Cell Sci *117*, 1567-1576.

Lunde, K., Belting, H.G., and Driever, W. (2004). Zebrafish pou5f1/pou2, homolog of mammalian Oct4, functions in the endoderm specification cascade. Curr Biol *14*, 48-55.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., *et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A *78*, 7634-7638.

Matoba, R., Niwa, H., Masui, S., Ohtsuka, S., Carter, M.G., Sharov, A.A., and Ko, M.S. (2006). Dissecting Oct3/4-regulated gene networks in embryonic stem cells by expression profiling. PLoS One *1*, e26.

Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T., and Yokota, T. (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. Embo J *18*, 4261-4269.

Matsui, Y., Zsebo, K., and Hogan, B.L. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell *70*, 841-847.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol *7*, 540-546.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Minucci, S., Botquin, V., Yeom, Y.I., Dey, A., Sylvester, I., Zand, D.J., Ohbo, K., Ozato, K., and Scholer, H.R. (1996). Retinoic acid-mediated down-regulation of Oct3/4 coincides with the loss of promoter occupancy in vivo. Embo J *15*, 888-899.

Mitchell, P.J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371-378.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.

Morrison, G.M., and Brickman, J.M. (2006). Conserved roles for Oct4 homologues in maintaining multipotency during early vertebrate development. Development *133*, 2011-2022.

Mullin, N.P., Yates, A., Rowe, A.J., Nijmeijer, B., Colby, D., Barlow, P.N., Walkinshaw, M.D., and Chambers, I. (2008). The pluripotency rheostat Nanog functions as a dimer. Biochem J *411*, 227-231.

Nakashima, K., Yanagisawa, M., Arakawa, H., Kimura, N., Hisatsune, T., Kawabata, M., Miyazono, K., and Taga, T. (1999). Synergistic signaling in fetal brain by STAT3-Smad1 complex bridged by p300. Science *284*, 479-482.

Nichols, J., Chambers, I., and Smith, A. (1994). Derivation of germline competent embryonic stem cells with a combination of interleukin-6 and soluble interleukin-6 receptor. Exp Cell Res *215*, 237-239.

Nichols, J., Chambers, I., Taga, T., and Smith, A. (2001). Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. Development *128*, 2333-2339.

Nichols, J., Davidson, D., Taga, T., Yoshida, K., Chambers, I., and Smith, A. (1996). Complementary tissue-specific expression of LIF and LIF-receptor mRNAs in early mouse embryogenesis. Mech Dev *57*, 123-131.

Nichols, J., Silva, J., Roode, M., and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development *136*, 3215-3222.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379-391.

Niehrs, C. (1999). Head in the WNT: the molecular nature of Spemann's head organizer. Trends Genet *15*, 314-319.

Nieuwkoop, P.D. (1997). Short historical survey of pattern formation in the endo-mesoderm and the neural anlage in the vertebrates: the role of vertical and planar inductive actions. Cell Mol Life Sci *53*, 305-318.

Nishimoto, M., Fukushima, A., Okuda, A., and Muramatsu, M. (1999). The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. Mol Cell Biol *19*, 5453-5465.

Nishimoto, M., Miyagi, S., Yamagishi, T., Sakaguchi, T., Niwa, H., Muramatsu, M., and Okuda, A. (2005). Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus. Mol Cell Biol *25*, 5084-5094.

Nishioka, N., Inoue, K., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R.O., Ogonuki, N., *et al.* (2009). The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. Dev Cell *16*, 398-410.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev *12*, 2048-2060.

Niwa, H., Masui, S., Chambers, I., Smith, A.G., and Miyazaki, J. (2002). Phenotypic complementation establishes requirements for specific POU domain and generic transactivation function of Oct-3/4 in embryonic stem cells. Molecular and cellular biology *22*, 1526-1536.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet *24*, 372-376.

Niwa, H., Sekita, Y., Tsend-Ayush, E., and Grutzner, F. (2008). Platypus Pou5f1 reveals the first steps in the evolution of trophectoderm differentiation and pluripotency in mammals. Evol Dev *10*, 671-682.

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell *123*, 917-929.

Norton, J.D. (2000). ID helix-loop-helix proteins in cell growth, differentiation and tumorigenesis. J Cell Sci *113 (Pt 22)*, 3897-3905.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., and Jenuwein, T. (2001). The polycomb-group gene Ezh2 is required for early mouse development. Mol Cell Biol *21*, 4330-4336.

Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D., and Niwa, H. (2006). Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. Biochem Biophys Res Commun *343*, 159-166.

Ohashi, Y., Brickman, J.M., Furman, E., Middleton, B., and Carey, M. (1994). Modulating the potency of an activator in a yeast in vitro transcription system. Mol Cell Biol *14*, 2731-2739.

Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M., and Hamada, H. (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell *60*, 461-472.

Okazawa, H., Okamoto, K., Ishino, F., Ishino-Kaneko, T., Takeda, S., Toyoda, Y., Muramatsu, M., and Hamada, H. (1991). The oct3 gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer. Embo J *10*, 2997-3005.

Okita, K., and Yamanaka, S. (2006). Intracellular signaling pathways regulating pluripotency of embryonic stem cells. Curr Stem Cell Res Ther *1*, 103-111.

Okumura-Nakanishi, S., Saito, M., Niwa, H., and Ishikawa, F. (2005). Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J Biol Chem *280*, 5307-5317.

Orkin, S.H., Wang, J., Kim, J., Chu, J., Rao, S., Theunissen, T.W., Shen, X., and Levasseur, D.N. (2008). The transcriptional network controlling pluripotency in ES cells. Cold Spring Harb Symp Quant Biol *73*, 195-202.

Paling, N.R., Wheadon, H., Bone, H.K., and Welham, M.J. (2004). Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling. J Biol Chem *279*, 48063-48070.

Palmieri, S.L., Peter, W., Hess, H., and Scholer, H.R. (1994). Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. Dev Biol *166*, 259-267.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell *1*, 299-312.

Pan, G.J., Chang, Z.Y., Scholer, H.R., and Pei, D. (2002). Stem cell pluripotency and transcription factor Oct4. Cell Res *12*, 321-329.

Park, S.W., Hu, X., Gupta, P., Lin, Y.P., Ha, S.G., and Wei, L.N. (2007). SUMOylation of Tr2 orphan receptor involves Pml and fine-tunes Oct4 expression in stem cells. Nat Struct Mol Biol *14*, 68-75.

Parvin, M.S., Okuyama, N., Inoue, F., Islam, M.E., Kawakami, A., Takeda, H., and Yamasu, K. (2008). Autoregulatory loop and retinoic acid repression regulate pou2/pou5f1 gene expression in the zebrafish embryonic brain. Dev Dyn *237*, 1373-1388.

Pasini, D., Bracken, A.P., Agger, K., Christensen, J., Hansen, K., Cloos, P.A., and Helin, K. (2008a). Regulation of stem cell differentiation by histone methyltransferases and demethylases. Cold Spring Harb Symp Quant Biol *73*, 253-263.

Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Mol Cell Biol *27*, 3769-3779.

Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. Embo J *23*, 4061-4071.

Pasini, D., Hansen, K.H., Christensen, J., Agger, K., Cloos, P.A., and Helin, K. (2008b). Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. Genes Dev *22*, 1345-1355.

Pesce, M., and Scholer, H.R. (2001). Oct-4: gatekeeper in the beginnings of mammalian development. Stem Cells *19*, 271-278.

Pietersen, A.M., and van Lohuizen, M. (2008). Stem cell regulation by polycomb repressors: postponing commitment. Curr Opin Cell Biol *20*, 201-207.

Qi, X., Li, T.G., Hao, J., Hu, J., Wang, J., Simmons, H., Miura, S., Mishina, Y., and Zhao, G.Q. (2004). BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. Proc Natl Acad Sci U S A *101*, 6027-6032.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425-432.

Reim, G., Mizoguchi, T., Stainier, D.Y., Kikuchi, Y., and Brand, M. (2004). The POU domain protein spg (pou2/Oct4) is essential for endoderm formation in cooperation with the HMG domain protein casanova. Dev Cell *6*, 91-101.

Remenyi, A., Lins, K., Nissen, L.J., Reinbold, R., Scholer, H.R., and Wilmanns, M. (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. Genes Dev *17*, 2048-2059.

Resnick, J.L., Bixler, L.S., Cheng, L., and Donovan, P.J. (1992). Long-term proliferation of mouse primordial germ cells in culture. Nature *359*, 550-551.

Riley, J.K., Carayannopoulos, M.O., Wyman, A.H., Chi, M., and Moley, K.H. (2006). Phosphatidylinositol 3-kinase activity is critical for glucose metabolism and embryo survival in murine blastocysts. J Biol Chem *281*, 6010-6019.

Riley, J.K., Carayannopoulos, M.O., Wyman, A.H., Chi, M., Ratajczak, C.K., and Moley, K.H. (2005). The PI3K/Akt pathway is present and functional in the preimplantation mouse embryo. Dev Biol *284*, 377-386.

Robertson, M. (1988). Homoeo boxes, POU proteins and the limits to promiscuity. Nature *336*, 522-524.

Rodda, D.J., Chew, J.L., Lim, L.H., Loh, Y.H., Wang, B., Ng, H.H., and Robson, P. (2005). Transcriptional regulation of nanog by OCT4 and SOX2. J Biol Chem *280*, 24731-24737.

Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirier, F., Rigby, P.W., and Staudt, L.M. (1990). A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature *345*, 686-692.

Ryan, A.K., and Rosenfeld, M.G. (1997). POU domain family values: flexibility, partnerships, and developmental codes. Genes Dev *11*, 1207-1225.

Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988). GAL4-VP16 is an unusually potent transcriptional activator. Nature *335*, 563-564.

Sato, N., Kondo, M., and Arai, K. (2006). The orphan nuclear receptor GCNF recruits DNA methyltransferase for Oct-3/4 silencing. Biochem Biophys Res Commun *344*, 845-851.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat Med *10*, 55-63.

Saxe, J.P., Tomilin, A., Scholer, H.R., Plath, K., and Huang, J. (2009). Post-translational regulation of Oct4 transcriptional activity. PLoS One *4*, e4467.

Schiemann, W.P., Bartoe, J.L., and Nathanson, N.M. (1997). Box 3-independent signaling mechanisms are involved in leukemia inhibitory factor receptor alpha- and gp130-mediated stimulation of mitogen-activated protein kinase. Evidence for participation of multiple signaling pathways which converge at Ras. J Biol Chem *272*, 16631-16636.

Scholer, H.R. (1991). Octamania: the POU factors in murine development. Trends Genet *7*, 323-329.

Scholer, H.R., Ciesiolka, T., and Gruss, P. (1991). A nexus between Oct-4 and E1A: implications for gene regulation in embryonic stem cells. Cell *66*, 291-304.

Scholer, H.R., Hatzopoulos, A.K., Balling, R., Suzuki, N., and Gruss, P. (1989). A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. Embo J *8*, 2543-2550.

Schoorlemmer, J., van Puijenbroek, A., van Den Eijnden, M., Jonk, L., Pals, C., and Kruijer, W. (1994). Characterization of a negative retinoic acid response element in the murine Oct4 promoter. Mol Cell Biol *14*, 1122-1136.

Schuldiner, M., Yanuka, O., Itskovitz-Eldor, J., Melton, D.A., and Benvenisty, N. (2000). Effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells. Proc Natl Acad Sci U S A *97*, 11307-11312.

Sears, R., Nuckolls, F., Haura, E., Taya, Y., Tamai, K., and Nevins, J.R. (2000). Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. Genes Dev *14*, 2501-2514.

Sekkai, D., Gruel, G., Herry, M., Moucadel, V., Constantinescu, S.N., Albagli, O., Tronik-Le Roux, D., Vainchenker, W., and Bennaceur-Griscelli, A. (2005). Microarray analysis of LIF/Stat3 transcriptional targets in embryonic stem cells. Stem Cells *23*, 1634-1642.

Sharov, A.A., Dudekula, D.B., and Ko, M.S. (2005). A web-based tool for principal component and significance analysis of microarray data. Bioinformatics *21*, 2548-2549.

Sharov, A.A., Masui, S., Sharova, L.V., Piao, Y., Aiba, K., Matoba, R., Xin, L., Niwa, H., and Ko, M.S. (2008). Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. BMC Genomics *9*, 269.

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS Biol *6*, e253.

Silva, J., Chambers, I., Pollard, S., and Smith, A. (2006). Nanog promotes transfer of pluripotency after cell fusion. Nature *441*, 997-1001.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722-737.

Singh, S.K., Kagalwala, M.N., Parker-Thornburg, J., Adams, H., and Majumder, S. (2008). REST maintains self-renewal and pluripotency of embryonic stem cells. Nature *453*, 223-227.

Singla, D.K., Schneider, D.J., LeWinter, M.M., and Sobel, B.E. (2006). wnt3a but not wnt11 supports self-renewal of embryonic stem cells. Biochem Biophys Res Commun *345*, 789-795.

Slack, J.M., Dale, L., and Smith, J.C. (1984). Analysis of embryonic induction by using cell lineage markers. Philos Trans R Soc Lond B Biol Sci *307*, 331-336.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature *336*, 688-690.

Smith, A.G., and Hooper, M.L. (1987). Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. Dev Biol *121*, 1-9.

Smith, S.T., and Jaynes, J.B. (1996). A conserved region of engrailed, shared among all en-, gsc-, Nk1-, Nk2- and msh-class homeoproteins, mediates active transcriptional repression in vivo. Development *122*, 3141-3150.

Smith, T.A., and Hooper, M.L. (1983). Medium conditioned by feeder cells inhibits the differentiation of embryonal carcinoma cultures. Exp Cell Res *145*, 458-462.

Snir, M., Ofir, R., Elias, S., and Frank, D. (2006). Xenopus laevis POU91 protein, an Oct3/4 homologue, regulates competence transitions from mesoderm to neural cell fates. Embo J *25*, 3664-3674.

Soodeen-Karamath, S., and Gibbins, A.M. (2001). Apparent absence of oct 3/4 from the chicken genome. Mol Reprod Dev *58*, 137-148.

Spaniol, P., Bornmann, C., Hauptmann, G., and Gerster, T. (1996). Class III POU genes of zebrafish are predominantly expressed in the central nervous system. Nucleic Acids Res *24*, 4874-4881.

Stahl, N., Farruggella, T.J., Boulton, T.G., Zhong, Z., Darnell, J.E., Jr., and Yancopoulos, G.D. (1995). Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. Science *267*, 1349-1353.

Stambolic, V., MacPherson, D., Sas, D., Lin, Y., Snow, B., Jang, Y., Benchimol, S., and Mak, T.W. (2001). Regulation of PTEN transcription by p53. Mol Cell *8*, 317-325.

Stavridis, M.P., Lunn, J.S., Collins, B.J., and Storey, K.G. (2007). A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. Development *134*, 2889-2894.

Stewart, C.L., Gadi, I., and Bhatt, H. (1994). Stem cells from primordial germ cells can reenter the germ line. Dev Biol *161*, 626-628.

Stewart, C.L., Kaspar, P., Brunet, L.J., Bhatt, H., Gadi, I., Kontgen, F., and Abbondanzo, S.J. (1992). Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. Nature *359*, 76-79.

Sturm, R.A., Das, G., and Herr, W. (1988). The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. Genes Dev *2*, 1582-1599.

Sturm, R.A., and Herr, W. (1988). The POU domain is a bipartite DNA-binding structure. Nature *336*, 601-604.

Sun, H., Lesche, R., Li, D.M., Liliental, J., Zhang, H., Gao, J., Gavrilova, N., Mueller, B., Liu, X., and Wu, H. (1999). PTEN modulates cell cycle progression and cell survival by regulating phosphatidylinositol 3,4,5,-trisphosphate and Akt/protein kinase B signaling pathway. Proc Natl Acad Sci U S A *96*, 6199-6204.

Surani, M.A., Hayashi, K., and Hajkova, P. (2007). Genetic and epigenetic regulators of pluripotency. Cell *128*, 747-762.

Sylvester, I., and Scholer, H.R. (1994). Regulation of the Oct-4 gene by nuclear receptors. Nucleic Acids Res *22*, 901-911.

Takahashi, K., Murakami, M., and Yamanaka, S. (2005). Role of the phosphoinositide 3-kinase pathway in mouse embryonic stem (ES) cells. Biochem Soc Trans *33*, 1522-1525.

Takahashi, K., Okita, K., Nakagawa, M., and Yamanaka, S. (2007a). Induction of pluripotent stem cells from fibroblast cultures. Nat Protoc *2*, 3081-3089.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007b). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Takeda, H., Matsuzaki, T., Oki, T., Miyagawa, T., and Amanuma, H. (1994). A novel POU domain gene, zebrafish pou2: expression and roles of two alternatively spliced twin products in early development. Genes Dev *8*, 45-59.

Tam, W.L., Lim, C.Y., Han, J., Zhang, J., Ang, Y.S., Ng, H.H., Yang, H., and Lim, B. (2008). T-cell factor 3 regulates embryonic stem cell pluripotency and self-renewal by the transcriptional control of multiple lineage pathways. Stem Cells *26*, 2019-2031.

Tanaka, S., Kunath, T., Hadjantonakis, A.K., Nagy, A., and Rossant, J. (1998). Promotion of trophoblast stem cell proliferation by FGF4. Science *282*, 2072-2075.

Tanaka, T.S., Kunath, T., Kimber, W.L., Jaradat, S.A., Stagg, C.A., Usuda, M., Yokota, T., Niwa, H., Rossant, J., and Ko, M.S. (2002). Gene expression profiling of embryo-derived stem cells reveals candidate genes associated with pluripotency and lineage specificity. Genome Res *12*, 1921-1928.

Tay, Y., Zhang, J., Thomson, A.M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature *455*, 1124-1128.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196-199.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.

Tolkunova, E.N., Fujioka, M., Kobayashi, M., Deka, D., and Jaynes, J.B. (1998). Two distinct types of repression domain in engrailed: one interacts with the groucho corepressor and is preferentially active on integrated target genes. Mol Cell Biol *18*, 2804-2814.

Triezenberg, S.J., Kingsbury, R.C., and McKnight, S.L. (1988a). Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. Genes Dev *2*, 718-729.

Triezenberg, S.J., LaMarco, K.L., and McKnight, S.L. (1988b). Evidence of DNA: protein interactions that mediate HSV-1 immediate early gene activation by VP16. Genes Dev *2*, 730-742.

Ura, H., Usuda, M., Kinoshita, K., Sun, C., Mori, K., Akagi, T., Matsuda, T., Koide, H., and Yokota, T. (2008). STAT3 and Oct-3/4 control histone modification through induction of Eed in embryonic stem cells. J Biol Chem *283*, 9713-9723.

Veeman, M.T., Axelrod, J.D., and Moon, R.T. (2003). A second canon. Functions and mechanisms of beta-catenin-independent Wnt signaling. Dev Cell *5*, 367-377.

Verrijzer, C.P., Strating, M., Mul, Y.M., and van der Vliet, P.C. (1992). POU domain transcription factors from different subclasses stimulate adenovirus DNA replication. Nucleic Acids Res *20*, 6369-6375.

Vigano, M.A., and Staudt, L.M. (1996). Transcriptional activation by Oct-3: evidence for a specific role of the POU-specific domain in mediating functional interaction with Oct-1. Nucleic Acids Res *24*, 2112-2118.

Wang, J., Levasseur, D.N., and Orkin, S.H. (2008). Requirement of Nanog dimerization for stem cell self-renewal and pluripotency. Proc Natl Acad Sci U S A *105*, 6326-6331.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364-368.

Wang, S.H., Tsai, M.S., Chiang, M.F., and Li, H. (2003). A novel NK-type homeobox gene, ENK (early embryo specific NK), preferentially expressed in embryonic stem cells. Gene Expr Patterns *3*, 99-103.

Ware, C.B., Horowitz, M.C., Renshaw, B.R., Hunt, J.S., Liggitt, D., Koblar, S.A., Gliniak, B.C., McKenna, H.J., Papayannopoulou, T., Thoma, B., *et al.* (1995). Targeted disruption of the low-affinity leukemia inhibitory factor receptor gene causes placental, skeletal, neural and metabolic defects and results in perinatal death. Development *121*, 1283-1299.

Wegner, M., Drolet, D.W., and Rosenfeld, M.G. (1993). POU-domain proteins: structure and function of developmental regulators. Curr Opin Cell Biol *5*, 488-498.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature *336*, 684-687.

Wood, H.B., and Episkopou, V. (1999). Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. Mech Dev *86*, 197-201.

Wu, D., Pang, Y., Ke, Y., Yu, J., He, Z., Tautz, L., Mustelin, T., Ding, S., Huang, Z., and Feng, G.S. (2009). A conserved mechanism for control of human and mouse embryonic stem cell pluripotency and differentiation by shp2 tyrosine phosphatase. PLoS One *4*, e4914.

Wu, Q., Chen, X., Zhang, J., Loh, Y.H., Low, T.Y., Zhang, W., Zhang, W., Sze, S.K., Lim, B., and Ng, H.H. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. J Biol Chem *281*, 24090-24094.

Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J.A., and Kosik, K.S. (2009). MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. Cell *137*, 647-658.

Xu, R.H., Peck, R.M., Li, D.S., Feng, X., Ludwig, T., and Thomson, J.A. (2005). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. Nat Methods *2*, 185-190.

Yates, A., and Chambers, I. (2005). The homeodomain protein Nanog and pluripotency in mouse embryonic stem cells. Biochem Soc Trans *33*, 1518-1521.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development *122*, 881-894.

Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281-292.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-523.

Yoshida, K., Chambers, I., Nichols, J., Smith, A., Saito, M., Yasukawa, K., Shoyab, M., Taga, T., and Kishimoto, T. (1994). Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. Mech Dev *45*, 163-171.

Yoshida-Koide, U., Matsuda, T., Saikawa, K., Nakanuma, Y., Yokota, T., Asashima, M., and Koide, H. (2004). Involvement of Ras in extraembryonic endoderm differentiation of embryonic stem cells. Biochem Biophys Res Commun *313*, 475-481.

Yoshikawa, Y., Fujimori, T., McMahon, A.P., and Takada, S. (1997). Evidence that absence of Wnt-3a signaling promotes neuralization instead of paraxial mesoderm development in the mouse. Dev Biol *183*, 234-242.

Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes Dev *9*, 2635-2645.

Zhang, J., Tam, W.L., Tong, G.Q., Wu, Q., Chan, H.Y., Soh, B.S., Lou, Y., Yang, J., Ma, Y., Chai, L., *et al.* (2006). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. Nat Cell Biol *8*, 1114-1123.

Zhang, Z., Liao, B., Xu, M., and Jin, Y. (2007). Post-translational modification of POU domain transcription factor Oct-4 by SUMO-1. Faseb J *21*, 3042-3051.

Zhong, X., and Jin, Y. (2009). Critical roles of coactivator p300 in mouse embryonic stem cell differentiation and Nanog expression. J Biol Chem *284*, 9168-9175.

Zhou, Q., Chipperfield, H., Melton, D.A., and Wong, W.H. (2007). A gene regulatory network in mouse embryonic stem cells. Proc Natl Acad Sci U S A *104*, 16438-16443.